உத்தமம்
INFITT



உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம்
International Forum for Information Technology in Tamil

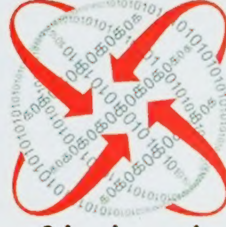


16 - ஆவது உலகத் தமிழ் இணைய மாநாடு
தொராண்டோ, கனடா

விழாமலர்

16 ஆவது உலகத் தமிழிணைய மாநாடு
தொராண்டோ, ஆகத்து 25-27, 2017

உத்தமம்
INFITT



உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம்

International Forum for Information Technology in Tamil



16 ஆவது உலகத் தமிழிணைய மாநாடு

தொராண்டோ, ஆகத்து 25-27 2017

தமிழ்த்தாய் வாழ்த்து

நீராரும் கடலுடுத்த நிலமடந்தைக் கெழிலொழுகும்
சீராரும் வதனமெனத் திகழ்பரதக் கண்டமிதில்
தெக்கணமும் அதிற்சிறந்த திராவிடநல் திருநாடும்
தக்கசிறு பிறைநுதலும் தரித்தநறும் திலகமுமே!
அத்திலக வாசனைபோல் அனைத்துலகும் இன்பமுற
எத்திசையும் புகழ்மணக்க இருந்தபெரும் தமிழணங்கே!

பல்லுயிரும் பலவுலகும் படைத்தளித்துத் துடைக்கினுமோர்
எல்லையறு பரம்பொருள்முன் இருந்தபடி இருப்பதுபோல்
கன்னடமுங் களிதெலுங்கும் கவின்மலையாளமும் துளுவும்
உன்னுதரத் தேயுதித்தே ஒன்றுபல வாகிடினும்
ஆரியம்போல் உலகவழக்கழிந் தொழிந்து சிதையாவுன்
சீரிளமைத் திறம்வியந்து செயன்மறந்து வாழ்த்துதுமே

தமிழ் மொழி வாழ்த்து

வாழ்க நிரந்தரம் வாழ்க தமிழ்மொழி
வாழிய வாழிய வே!
வான மளந்த தனைத்தும் அளந்திடும்
வண்மொழி வாழிய வே!
ஏழ்கடல் வைப்பினுந் தன்மணம் வீசி
இசைகொண்டு வாழிய வே!
எங்கள் தமிழ்மொழி! எங்கள் தமிழ்மொழி!
என்றென்றும் வாழிய வே!
சூழ்கலி நீங்கத் தமிழ்மொழி ஓங்கத்
துலங்குக வையக மே!
தொல்லை வினைதரு தொல்லை யகன்று
சுடர்க தமிழ்நா டே!
வாழ்க தமிழ்மொழி! வாழ்க தமிழ்மொழி!
வாழ்க தமிழ்மொழி யே!
வானம் அறிந்த தனைத்தும் அறிந்து
வளர்மொழி வாழிய வே!

உள்ளடக்கம்

வாழ்த்துரைகள்:

கட்டுரைகள்:

உலகத் தமிழிணைய மாநாடு – 2017
நிகழ்ச்சி நிரல்

முதலாம் நாள் : வெள்ளிக்கிழமை (25 ஆகத்து)

08.15 – 09.00 வருகைப் பதிவும் காலைச் சிற்றுண்டியும்
09.00 – 11.00 விழா தொடக்க நிகழ்ச்சிகள்
மங்கல விளக்கேற்றல்
தமிழ்த்தாய் வாழ்த்து
கனேடிய நாட்டுப்பண்
வரவேற்பு நடனம்
வரவேற்புரை : பேராசிரியர். செ. இரா. செல்வக்குமார்
பேராசிரியர் கு. பொன்னம்பலம்
தலைமை விருந்தினர் உரை: பேராசிரியர். சு. பசுபதி
சிறப்புரை: பேராசிரியர் மு. பொன்னவைக்கோ
சிறப்பு விருந்தினர் உரைகள்:
1) மரு. சம்பந்தன் (Harvard Tamil Chair donor)
2) திரு. அ. முத்துலிங்கம் (Harvard Tamil Chair donor)
3) திரு. சிவன் இளங்கோ (VP, Candian Tamil Congress)
4) முனைவர். திருமதி. செல்வநாயகி ஸ்ரீதாஸ் (Annamalai Canada Campus)

11.00 – 11.30 தேநீர் இடைவேளை
11.30 – 12.30 சிறப்பு விரிவுரை – ஆன்றூ வாங்கு Andrew Wong – [miw IC 220]
12.30 – 13.30 பகல் உணவு இடைவேளை
13.30 – 15.15 அமர்வு 1A (5 x20 நிமிடம்) – [அறை IC 220]
அமர்வு IB (5 x 20 நிமிடம்) – [அறை IC 230]
15.15 – 15.30 தேநீர் இடைவேளை
15.30 – 17.00 அமர்வு 2 (பயிற்சிப் பட்டறை 3 x 30 நிமிடம்) – [அறை IC 220]

இரண்டாம் நாள்: சனிக்கிழமை (26 ஆகத்து)

08.30 - 09.00 காலைச் சிற்றுண்டி
09.00 - 10.30 அமர்வு 3 (மெய்நிகர் 4 x 20 நிமிடம்) - [அறை IC 130]
1030 - 10.45 தேநீர் இடைவேளை
10.45 - 11.45 சிறப்பு விரிவுரை - முத்து நெடுமாறன் Muthu Nedumaran
11.45 - 12.30 அமர்வு 4A (மெய்நிகர், 2 x 20 நிமிடம்) – [அறை IC 130]
அமர்வு 4B (2 x 20 நிமிடம்) – [அறை IC 220]
12.30 - 13.30 பகல் உணவு இடைவேளை
13.30 - 15.15 அமர்வு 5 (5 x 20 நிமிடம்) – [அறை IC 130]
15.15 - 15.30 தேநீர் இடைவேளை
15.30 - 17.00 அமர்வு 6 (பயிற்சிப் பட்டறை 60 நிமிடம் 30 நிமிடம்) – [அறை IC 130]

மூன்றாம் நாள்: ஞாயிற்றுக்கிழமை (27 ஆகத்து)

08.30 - 09.00 காலைச் சிற்றுண்டி
09.00 - 10.30 அமர்வு 7 (மெய்நிகர் 4 x 20 நிமிடம்) – [அறை இல. IC 130]
10.30 - 10.45 தேநீர் இடைவேளை
10.45 - 11.45 சிறப்பு விரிவுரை - கு. கல்யாணசுந்தரம் K. Kalyanasundaram
11.45 - 12.30 அமர்வு 8A (மெய்நிகர் 2x20 நிமிடம்) - [அறை இல. IC 130]
அமர்வு 8B (2 x 20 நிமிடம்) - [அறை இல. IC 220]
12.30 - 13.30 பகல் உணவு இடைவேளை
13.30 - 15.00 விழா நிறைவு நிகழ்ச்சிகள்
சிறப்புரை - முனைவர் மு. பொன்னவைக்கோ
பரிசு வழங்கல்
கருத்துரைகள்
நன்றியுரை

Keynote Speakers



Muthu Nedumaran
(Founder, CEO, Murasu Inc)



Andrew K. C Wong
(Distinguished Professor Emeritus)



Dr. K. Kalyanasundram
(Swiss Federal Institute of Technology [EPFL])

பயிற்சிப்பட்டறை பற்றிய விபரங்கள்

1) Speech Recognition by A. Madhavaraj, Indian Institute of Science, Bangalore, India

2) எழில் மொழிப் பட்டறை – தமிழில் நிரல் எழுதலாம் – முத்தையா அண்ணாமலை, அமெரிக்கா

3) Thirukkural Bot - by Shiva Subramaniam, United Kingdom

4) Digital Scholarship: Tools and Techniques - Kim Pham, Toronto

5) A Walk with Big Data - by Muthu Arumugam, Kogentix, Chicago

UNIVERSITY OF WATERLOO

OFFICE OF THE PRESIDENT
519-888-4400
president@uwaterloo.ca | uwaterloo.ca

August 25, 2017

Greetings,

Welcome to Canada for the 16th Tamil Internet Conference.

I would like to thank the International Forum for Information Technology in Tamil for bringing this important event to Canada - and for having the foresight to plan this for the year when we are celebrating our country's 150th anniversary.

The University of Waterloo Centre for Pattern Recognition and Machine Intelligence is proud to co-sponsor this event, along with IEEE Canada. As well, I'm personally thrilled to see that two Waterloo professors have taken on leadership roles, with Prof. C.R. (Selva) Selvakumar serving as Conference Chair and Prof. Kumaraswamy Ponnambalam working as Chair of the Local Organizing Committee.

We are Canada's top innovation university. With more than 36,000 students, Waterloo is home to the world's largest co-operative education system of its kind. Our unmatched entrepreneurial culture, combined with an intensive focus on research, powers one of the top innovation hubs in the world. That makes Waterloo a natural sponsor for this conference.

Our University is at the centre of outstanding research being conducted in the Waterloo Region Innovation Ecosystem. The Toronto-Waterloo Corridor is one of the largest technology hubs in the world with 15,000 tech companies, 5,200 tech startups, and 200,000 tech workers.

With a focus on Deep Learning and Data Science, you are discussing timely subjects. I know that this conference will make an important contribution to a growing field.

I wish you all the best.

Sincerely,

Feridun Hamdullahpur
President and Vice-Chancellor



200 UNIVERSITY AVENUE WEST, WATERLOO, ON, CANADA N2L 3G1



Historical & Cultural Studies
UNIVERSITY OF TORONTO
SCARBOROUGH

Greetings, Vanakkam.

It is my honor and pleasure to welcome you all to the 16th Tamil Internet Conference. I would like to thank the International Forum for Information Technology in Tamil, led by Prof. Selvakumar of the University of Waterloo, for bringing this important conference to Toronto and particularly to the University of Toronto-Scarborough Campus. Tamil Worlds Initiative at UTSC is delighted to support this event and we hope you will all have a productive meeting.

UTSC□s Tamil Worlds Initiativewas established in 2013 to promote excellence in public programing, scholarly research and teaching of Tamil-related issues at the University of Toronto. The initiative is based in the Department of Historical and Cultural Studies at the University of Toronto-Scarborough and we try very much to engage with the multiplicity, dynamism, and global spread of Tamil-speaking groups. I am personally very excited that among the topics to be discussed at this meeting are open-source Tamil software, digital technology for learning Tamil, and the Tamil digital humanities. These are all topics that need to be foregrounded as the digital humanities are important to the future of Tamil-based knowledge and cultural memory. We hope this is the beginning of a collaboration on these issues and I look forward to working together.

My very best wishes for your discussions,

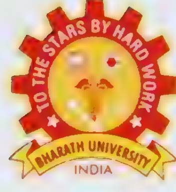
Bhavani Raman

Bhavani Raman

Chair,
Tamil Worlds Initiative Program Committee,
Historical and Cultural Studies, UTSC.

1265 Military Trail, Toronto, ON M1C 1A4 Canada
http://www.utsc.utoronto.ca/~hcs/

Padmashri
Dr.AVVAI NATARAJAN
CHANCELLOR - BHARATH UNIVERSITY
FORMER VICE-CHANCELLOR
FORMER SECY TO GOVT. OF TAMILNADU (RETD.)



OFFICE :
30, B.B.C. ORCHID APARTMENTS
TILAK STREET, THEAGARAYA NAGAR,
CHENNAI - 600 017.
98410 19950 | 044-28340020
thamizhavvai @ hotmail.com

கனடா மநாடு கரம் நீட்டுமா?

16.08.2017

அளப்பரிய அறிவின் வளப்பெரும் திலகமாகத் திகழ்வது இணையத்தின் எழுச்சியாகும். உலக நாடுகளில் சேர்ந்தாரை அரவணைக்கும் செழிப்பும் சிறப்பும் கொண்ட ஒப்பரிய நாடு என்ற பெருமை கனடாவுக்குத்தான் உரியது.

எண்ணும் எழுத்தும் கண் எனத் தகும் என்பது மாறிக் கணினியும் இணையமுமே நமக்கு வாய்த்த இரு கண்கள் என்ற விரைந்த வளர்ச்சியில் உலக நாடுகள் சிறந்து மிளிர்கின்றன. காலத்தால் கண்டறிய முடியாத தொன்மை கொண்ட மொழி தமிழ்மொழிதான் என்று நெஞ்சு உயர்த்திப் பேசினாலும் அறிவியல் வளர்ச்சியினுடைய ஆக்கங்கள் இன்னும் முழுமையாகத் தமிழில் தோயவில்லை.

இந்நிலையில், கனடா நாட்டில் வாழும் தமிழர்கள், தாம் வாழும் நாட்டுக்கு நன்றியுடைவர்களாக சிறப்படைவதோடு நூற்றி ஐம்பதாவது ஆண்டு விடுதலைத் திருநாளை எண்ணி நெகிழ்வதும் கணினி உயர்வால் தமிழை உயர்த்தும் மகிழ்வும் கொண்டு உலக இணைய மாநாட்டை நடத்துவது பொருத்தமாகும்.

தமிழில் இணைய வளர்ச்சியைப் பற்றி கருதும்போது, கனடா நாட்டு இளைஞர்கள் தமிழகத்திற்குத் திரண்டு வந்து தமிழர்களுக்கு நடுவில் இந்த இணையக் கருத்துகளையெல்லாம் வலியுறுத்த வேண்டுமென்பது என் பெருவிருப்பமாகும்.

பொறியியல் கல்வி தேய்ந்து சாம்பலாகிப் போகும் மன வருத்தத்தில் தமிழக மாணவர்கள் உள்ளனர். பொறியியல் கல்லூரிகள் பல புகை சூழ்ந்த இடிந்த மாடங்களாகி வருகின்றன. இந்தக் கொடுமையை நீங்குவது பற்றியும் மாநாடு கருதும் என்று எதிர்பார்க்கிறேன்.

யார் யாரை வாழ்த்துவது?

அன்புள்ள,

ஔவை நடராசன்



பதினாறாவது தமிழ் இணைய மாநாட்டிற்கு வாழ்த்து

முனைவர் மு.பொன்னவைக்கோ,
இணைவேந்தர், பாரத் பல்கலைக்கழகம், சென்னை

முதல் தமிழ்இணைய மாநாடு, 1997–ஆம் ஆண்டு, மே மாதம் சிங்கப்பூரில் 'தமிழ் இணையம் 97' என்ற பெயரில் நடைபெற்றது. Internet என்னும் ஆங்கில சொல்லுக்கு 'இணையம்' என்னும் தமிழ்ச்சொல்லை கணித்தமிழ் உலகிற்கு வழங்கியதும் இந்த மாநாட்டில்தான். இரண்டாவது தமிழ்இணைய மாநாடு, 'தமிழ் இணையம் 99' என்ற பெயரில் சென்னையில் 1999–ஆம் ஆண்டு பிப்ரவரி மாதம் நடைபெற்றது. இந்த மாநாட்டில் எடுத்தமுடிவின்படி தமிழ் இணையப்பல்கலைக்கழகம் தோன்றியது. இம்மாநாட்டை அடுத்து 'தமிழ் இணையம் 2000' மாநாடு இலங்கையில் நடத்தத் திட்டமிடப்பட்டு இருந்தது. அந்த மாநாட்டின் முன்னேற்பாட்டுக் கூட்டம் 2000–ஆம் ஆண்டு நவம்பர் மாதம் இலங்கையில் நடைபெற்றது. அந்தக் கூட்டத்தில் சுவிட்சர்லாந்து நாட்டிலிலுள்ள முனைவர் கு.கல்யாணசுந்தரம் அவர்கள், உலகஅளவில் ஒரு இணையத்தமிழ் ஆய்வுக்குழு அமைக்க வேண்டுமென வரைந்து அனுப்பியிருந்த திட்டத்தை திரு.அருண்மகிழ்நன் முன்மொழிந்தார். அந்தக் கலந்துரையாடலில் பிறந்ததுதான் 'உத்தமம்' என்னும் உலகத்தமிழ்த் தகவல் தொழில் நுட்பமன்றம். இம்மன்றத்திற்குப் பெயரிடும் பெருமை எனக்குக் கிட்டியது நான் பெற்ற பேறு. அந்த மாநாட்டை அடுத்து சிங்கப்பூரில் 2000–ஆம் ஆண்டு ஜூலை மாதம் 22–24–ஆம் நாட்களில் நடைபெற்ற தமிழ் இணைய மாநாட்டில் 'உத்தமம்' (INFITT) தொடங்கி வைக்கப்பட்டது. இணையத்தமிழின் ஆய்விற்காக உத்தமத்தில் தமிழ்க் கலைச்சொல் தொகுப்பு, யூனிகோடு தமிழ் (UNICODE Tamil) ஆய்வு, இணையதள தமிழ் முகவரி வடிவமைத்தல், தமிழ் வரிவடிவக் குறியீட்டுத் தரப்பாடு, ஆங்கில வரிவடிவத் தமிழ்த் தரப்பாடு, தமிழ் எழுத்துரு படித்து அறிதல் (Tamil OCR), லினக்ஸில் தமிழ் (Tamil in Linux), தமிழ் அனைத்து எழுத்துரு 16–பிட்டு தரப்பாடு (TACE 16) ஆகிய பணிகளுக்காக எட்டு ஆய்வுப் பணிக்குழுக்கள் (Working Groups) நிறுவப்பட்டன. பணிக்குழுக்களின் செயற்பாடுகள் பற்றியும் கணித்தமிழ் மற்றும் இணையத்தமிழ் வளர்ச்சி பற்றியும் ஒவ்வொரு ஆண்டும் உத்தமம் நடத்திய தமிழ் இணைய மாநாடுகளில் கலந்தாய்வு செய்யப்பட்டுள்ளன. இதுவரை (1997–லிருந்து 2016–வரை) சிங்கப்பூர், தமிழ்நாடு,, மலேசியா, அமெரிக்கா, ஜெர்மனி, புதுச்சேரி ஆகிய நாடுகளில் 15 தமிழ் இணைய மாநாடுகள் நடைபெற்றுள்ளன. இப்பொழுது 16–வது மாநாடு கனடாவில் தொராண்டோ நகரில் நிகழவுள்ளது அறிய மிகுந்த மகிழ்ச்சி அளிக்கின்றது.

இந்த 20 ஆண்டுகளில் நாம் சாதித்தவை என்ன என்று எண்ணிப் பார்க்க வேண்டிய நேரம் இது. ஏராளமான தமிழ் இணையதளங்கள் பிறந்துள்ளன. பல்வேறு எழுத்துரு தரப்பாடுகளிலிருந்து இரண்டு எழுத்துருத்தரப்பாடுகள் – ஒருங்குறி தமிழ் (Unicode Tamil), அனைத்து எழுத்துருத்தரப்பாடு (TACE 16), ஆகிய இரண்டு தரப்பாடுகள் தமிழக அரசால் அரசின் தரப்பாடுகளாக ஏற்கப் பெற்றுள்ளோம். சொற்செயலிகள், தமிழ் எழுத்துரு அறி மென்மம் (Tamil OCR), பேச்சுத்தமிழை எழுத்துத்தமிழாக்கும் மென்மம், எழுத்துத்தமிழை பேச்சுத்தமிழாக்கும் மென்மம், தமிழ் இயல் மொழிச்செயலாக்கம் போன்ற பல்வேறு மென்பொருள் உருவாக்கப் பெற்றுள்ளோம். உலகு தழுவி வாழும் தமிழ் மக்களும், தமிழில் ஈடுபாடு உள்ள மற்றையோரும், தமிழ் மொழியைக் கற்கவும், தமிழர் வரவாறு, கலை, இலக்கியம், பண்பாடு பற்றி அறிந்து கொள்ளவும், மழலைக்கல்வி முதல் பட்டப் படிப்பிற்கான பாடப் பொருள்களையும்,

மிகப்பெரிய தமிழ் மின்நூலகத்தையும் தன்னுட்கொண்டு தமிழ்ப்பணி ஆற்றி வரும் சிறந்ததொரு தமிழ் இணையப் பல்கலைக்கழகம் கிடைக்கப் பெற்றுள்ளோம். இங்கே குறிப்பிட்ட படைப்புகளெல்லாம் மக்களைப் போய்ச் சேர்ந்துள்ளனவா? தமிழை ஆட்சி மொழியாகவும் வழக்கு மொழியாகவும் கொண்டுள்ள நாடுகளின் அரசுக்கள் ஏற்று செயல் படுத்துகின்றனவா? இல்லையெனில் அதற்கு உத்தமம் என்ன செய்யவேண்டும்? எப்படிச் செயல்படவேண்டும்? என்பவற்றை இம்மாநாட்டில் கலந்தாய்ந்து முடிவு செய்யவேண்டிய நிலையில் உள்ளோம். இணையப் பயன்பாட்டிற்கு ஒருங்குறி தமிழ் எழுத்துரு (Unicode Tamil) தரப்பாடு என்றும், பிற பயன்பாடுகளுக்கு அனைத்து எழுத்துரு (TACE-16) தரப்பாடு என்றும் தமிழக அரசு அறிவித்துள்ளது. எல்லாப் பன்பாடுகளுக்கும் அனைத்து எழுத்துரு (TACE 16) தரப்பாடே சிறந்தது எனபதை பல்வேறு ஆய்வுகள் வெளிப்படுத்தியுள்ளன. இந்த அனைத்து எழுத்துரு (TACE 16) தரப்பாட்டின் பயன்பாடு கூடினால், ஒருங்குறி சேர்த்தியம் (Unicode consortium) ஒருங்குறி தளத்தில் 32–பிட்டு அமைப்பில் இந்த அனைத்து எழுத்துரு (TACE-16) தரப்பாட்டினை சேர்க்க இசைவளித்துள்ளது. எனவே, அனைத்து எழுத்துரு (TACE 16) தரப்பாட்டினை பல்வேறு பயன்பாடுகளில் செயல்படுத்தி அதன் பயன்பாட்டினை பெருக்குமாறு கணித்தமிழ் அன்பர்களையெல்லாம் அன்புடன் கேட்டுக் கொள்கின்றேன். இன்று ஜப்பான், கொரியா போன்ற நாடுகளில் செயல்படும் கணிப்பொறிகளுக்கு ஆங்கிலம் தெரியாது. ஜப்பான், கொரிய மொழிகளில் கொடுக்கப்படும் கட்டளைகளை மட்டுமே புரிந்நதுகொண்டு செயல்படுகின்றன. உலக மொழிகளின் தாய்மொழியாகிய தமிழ் மொழிக் கட்டளைகளால் இயங்கும் கணிப்பொறியை வடிவமைக்க இயலாதா என்ன? அனைத்து எழுத்துரு (TACE-16) தரப்பாடு செய்யப்படும்வரை இயலா நிலை இருந்தது. இப்பொழுது அத்தடை இல்லை. இனி அப்படியொரு கணிப்பொறியைக் காண்பது எப்போழ்து? Assembler போன்ற அமைப்புச் செயல்மொழி (System software) windows போன்ற இயக்க மென்பொருள் (Operating System) ஆகியவற்றை அனைத்து எழுத்துரு (TACE-16) தரப்பாட்டில் வடிவமைத்து ஒரு முழுமையான தமிழ்க்கணினியை படைக்க முடியும். இப்படியொரு முழுமையான தமிழ்க்கணினியை படைத்து வழங்குபவருக்கு உருபா ஒரு இலக்கம் பரிசு வழங்கப்படும் என்று மலேசியாவில் நடைபெற்ற 12–வது தமிழ் இணைய மாநாட்டின் போழ்தே அறிவித்திருந்தேன். ஆனால் இன்றுவரை அது நிகழவில்லை. விரைவில் அப்படியொரு முழுமையான தமிழ்க்கணினி பயன்பாட்டிற்க்கு வரவேண்டும் என்பது எனது பேரவா? என் கனவு நிறைவேறுமா? இப்படி பல்வேறு கணித்தமிழ் ஆக்கப்பணிகள் நம்முன் கிடக்கும்பொழுது அவற்றில் கவனத்தைச் செலுத்துவதற்குப் பதிலாக நம்முள் ஒற்றுமையின்றி பல குழுக்களாகப்பிரிந்து போராடிக் கொண்டுளோம். இதனால் நாம் இழந்தவை பல. அண்மையில் நிகழ்ந்த பல செயற்பாடுகள் வேதனையையும் உறுப்பினர்களுக்கிடையே மனக்கசப்பையும் ஏற்படுத்தியுள்ளன. நாம் அனைவரும் ஒற்றுமையாக செயல்பட்டாலொழிய உத்தம அமைப்பில் சாதனைகளைப் படைக்க முடியாது. ஒன்றுபடுங்கள் வென்று வாகை சூடுங்கள் என்று உத்தம அமைப்பின் உறுப்பினர்கள் அனைவரையும் அன்போடு கேட்டுக்கொள்கின்றேன். வேற்றுமைகளை மறந்து பதினாறாவது தமிழ் இணைய மாநாட்டு சிறப்பாக நடைபெற அனைவரும் ஒத்துழைக்க வேண்டும் என்று எனது உளங்கனிந்த வாழ்த்துக்களை தெரிவித்துக் கொள்கின்றேன்.

வாழ்க தமிழ்! வளர்க கணித்தமிழ்!

அன்புடன்

மு.பொன்னவைக்கோ

முனைவர் **க.பாஸ்கரன்**
துணைவேந்தர்

மின் அஞ்சல்: rg.bhaskaran@gmail.com
இணையதளம் : www.tamiluniversity.ac.in



தமிழ்ப் பல்கலைக்கழகம்
தஞ்சாவூர்–613010 – தமிழ்நாடு, இந்தியா
அலுவலகம் : 04362–227040
இல்லம் : 04362–226741
நிகரி : 04362 226159

நாள்: 11.08.2017

வாழ்த்துரை

உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றத்தின் 16 ஆவது இணைய மாநாடு கனடநாட்டுத் தொராண்டோ பல்கலைக்கழகத்தில் நடைபெறுவது அறிந்து பெரிதும் மகிழ்ச்சியடைகிறேன். இம் மாநாட்டில் "ஆழ்தமிழின் அழகும் ஆழக்கற்றல் திறனும்" என்பதும் "தமிழில் தரவு அறிவியல்" என்பதும் மாநாட்டு முழக்கங்களாக முன்வைக்கப்பட்டுள்ளன. இது பெரிதும் பாராட்டுதற்குரியது.

உலக முதல் மொழியாம் தமிழ் மொழி இன்று பல்வேறு அறிவியல் தொழில்நுட்பங்களுடன் இயங்கிவருகிறது. தமிழை "என்றுமுள தென்தமிழாக" இருக்கச்செய்தற்குரிய முன்னெடுப்புகளில் கணினித் தமிழும் ஒன்றாகச் செயல்பட்டுவருகிறது. கணினித்தமிழ் தமிழ்மொழியைப் பரவலாக்கம் செய்வதையும் தாண்டி உலகமக்களின் இதயங்களில் எல்லாம் உயர்தனிச் செம்மொழியின் பண்பாட்டுக் கூறுகளை ஆழப் பதியச்செய்து வருகிறது. உலகியல் போக்கினை உணர்ந்து தமிழ்மொழியை வளர்த்தெடுக்கும் இணையத் தமிழ் வழியாக உலகமொழிகளின் அறிவையும் ஆற்றலையும் தமிழ்மொழியாளர்களுக்குக் கொண்டுசேர்க்க வேண்டும்.

இயற்கைமொழியாய் இருந்த பேச்சுத் தமிழுக்கு எழுத்துவடிவம் தோன்றிய பொழுதே இத்தமிழ் மொழியின் தொழில்நுட்ப வளர்ச்சி தொடங்கிவிட்டது. எழுதப்பட்டவைகளுக்கும் எழுதப்படாதவைகளுக்குமான இலக்கணங்கள் தோன்றிய பொழுதே தமிழ்மொழிக்கான பாதுகாப்பும் தொடங்கிவிட்டது. மொழிக்கான வளர்ச்சிப் போக்கில் கணினி, இணையம் எனும் அறிவியல் தொழில்நுட்ப வளர்ச்சிகளைத் தமிழ் மொழி இன்று தன்வசமாக்கி வளர்ந்துவருகிறது. கணினியில் தமிழ்மொழியைக் காணமுடியாதா என்று ஏங்கியிருந்த காலங்கள் கடந்துவிட்டன. கணினியில் தமிழ் காட்சிதந்து மகிழ்ச்சியளிக்கிறது. இன்று தானியங்கிச் சொற்பிழைத் திருந்தி மற்றும் இலக்கணப் பிழைத்திருத்திகள் முதலிய தொழில்நுட்ப வளர்ச்சிகள் தமிழுக்கும் வந்துவிட்டன.

ஆய்வேடுகள் இன்று கணினிவழித் தரப்படுத்தப்பட்டு வருகின்றன. ஆய்வுகளில் பிற ஆய்வேடுகளைப் படியெடுத்துப் பயன்படுத்தும் தவறான போக்குகளைத் தடுக்கும் வகையில் ஆய்வுநெறிமுறைகள் கணினிவழி செயல்படுத்தப்பட்டுவருகின்றன. ஆசிரியர் மாணவர் வருகைப்

பதிவுகள் இன்று மின்வழிப்பதிவுகளாக மாற்றம் பெற்றுவிட்டன. இவையாவும் மொழி வளர்ச்சிக்கும் தரமான ஆய்விற்கும் பெரிதும் துணைநிற்பன என்பதில் ஐயமில்லை.

இன்று எண்ணற்ற மின் நூலகங்கள் வந்துவிட்டன. நூலக அடுக்குகளில் நுழைந்து புத்தகங்களைத் தேடிய காலங்கள் கடந்துவருகின்றன. கணினிவழியில் உலகில் உள்ள எல்லா நூலக இருப்புகளையும் மாணவர்கள் அறிந்துகொள்ள முடிகிறது. அந்நூல்களைத் தரவிறக்கம் செய்தோ அல்லது அவற்றை அஞ்சல் வழியில் பெற்றோ பயன்படுத்த முடிகிறது. இவ்வளர்ச்சி மாணவர்களின் சிந்தனைத் திறனை மேலும் வளப்படுத்துவதற்கும் ஆய்வுக்காலங்களைச் சிக்கனப்படுத்துவதற்கும் துணைநிற்கிறது.

கணினித் தமிழின் வளர்ச்சிகள் வியந்துபார்க்கும் அளவில் அமைந்துள்ளதற்கு ஏற்ப அவ்வளர்ச்சியைப் பயன்படுத்துவோர் எண்ணிக்கையும் அமைய வேண்டும். கணினி பொழுதுபோக்கு ஊடகம் அல்ல. அது நம் பொழுதுகளை நல்வழியில் பொன்மயமாக்கும் ஊடகங்களாக வளரவேண்டும். அனைத்துத் துறை சார்ந்த நூல்களும் படிப்புகளும் தமிழ்வழியில் அமைதற்குக் கணினித் தமிழ் வழிவகை செய்யவேண்டும். எம்மொழியாளனுக்கும் தமிழ்மொழியாளன் குறைந்தவன் அல்லன். எவ்வித தொழில்நுட்ப வளர்ச்சிக்கும் தமிழ்மொழியின் ஆற்றலும் குறைந்ததல்ல என்னும் நிலையை உருவாக்க வேண்டும். அதற்கு இதுபோன்ற மாநாடுகள் தொடர்ந்து நடைபெறவேண்டும். இவ்வாறான மாநாடுகளை நடத்துவதற்கு உலகத் தமிழ் அமைப்புகளும் கல்விசார் நிறுவனங்களும் ஆர்வலர்களும் துணைநிற்க வேண்டும். தமிழை வளர்த்தெடுக்கும் இம்மாநாட்டின் முழக்கங்கள் வெற்றிபெற வாழ்த்துகிறேன்.

(முனைவர் க.பாஸ்கரன்)



මනෝ ගනේසන් ජාතික සහජීවනය, සංවාද හා රාජ්‍ය භාෂා අමාත්‍ය	**மனோ கணேசன்** தேசிய சகவாழ்வு, கலந்துரையாடல் மற்றும் அரசகரும மொழிகள் அமைச்சர்	**MANO GANESAN** Minister of National Co-existance, Dialogue and Official Languages

මගේ අංකය எனது இல. My No.	ඔබේ අංකය உமது இல. Your No.	දිනය திகதி Date

தேசிய சகவாழ்வு, கலந்துரையாடல் மற்றும் அரசகரும மொழிகள் அமைச்சரின் வாழ்த்துச் செய்தி

உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம் (உத்தமம்) ஏற்பாடு செய்துள்ள 16 ஆவது தமிழ் இணைய மாநாட்டு சிறப்பு மலர் வெளியீட்டிற்கு வாழ்த்து செய்தியினை தெரிவித்துக் கொள்வதில் மகிழ்ச்சியடைகின்றேன்.

1997 ஆம் ஆண்டு முதல் உலகின் வெவ்வேறு நாடுகளில் தமிழ் இணைய மாநாட்டினை நடத்தி தமிழ் தொழில் முனைவோர், தமிழ்க் கணினியாளர்கள், நிரலர்கள், பேராசிரியர்கள், ஆசிரியர்கள், மாணவர்களை தமிழால் ஒன்றிணைக்க முயல்வது "உத்தமம்" நிறுவனத்தின் சிறப்பாகும்.

அமெரிக்கா நாட்டில் பதிவு செய்யப்பட்டு, ஐரோப்பிய நாடுகளில் உறுப்பினர்களை கொண்டு மின்னுட்ப வல்லுநர்களும், மொழியியல் அறிஞர்களும், கல்வியாளர்களும், ஒன்றுகூடும் இந்த மாநாடு தமிழ் மொழி பற்றினால் பல துறையினரையும் ஒன்றிணைக்கும் ஒரு சிறந்த தளமாக உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம் திகழ்கின்றது.

ஆழ்தமிழின் அழகும் ஆழக்கற்றல் திறனும், தமிழில் தரவு அறிவியல் என்ற இரண்டு கருத்துக்களை முன் வைத்து ஒழுங்கு செய்யப்பட்டுள்ள இந்த மாநாட்டு கருத்தரங்கினை ஏற்பாடு செய்த மாநாட்டு ஏற்பாட்டுக் குழுவினருக்கு எனது மனமார்ந்த வாழ்த்துக்கள்.

கொழும்பு மாவட்ட பாராளுமன்ற உறுப்பினர் என்ற வகையிலும், தேசிய சகவாழ்வை முன்னெடுக்கும் தேசிய சகவாழ்வு, கலந்துரையாடல் மற்றும் அரசகரும மொழிகள் அமைச்சர் என்ற வகையிலும் இந் நிகழ்வு இனிதே நடைபெற வாழ்த்தினை தெரிவித்துக் கொள்கின்றேன்.

நன்றி.

மனோ கணேசன் பா.உ.
தேசிய சகவாழ்வு கலந்துரையாடல் மற்று.
அரசகரும மொழிகள் அமைச்சர்
#40, புத்கமுவ வீதி.
இராஜகிரிய.

மனோ கணேசன்
கொழும்பு மாவட்ட நாடாளுமன்ற உறுப்பினர்
தேசிய சகவாழ்வு, கலந்துரையாடல் மற்றும் அரசகரும மொழிகள் அமைச்சர்

தலைவர் - ஜனநாயக மக்கள் முன்னணி
தலைவர் - தமிழ் முற்போக்கு கூட்டணி

திகதி	:	14-08-2017
கைப்பேசி	:	(+94)777 312770
மின்னஞ்சல்	:	leader@dpflanka.org



தலைவரின் வாழ்த்துரை

செல்வமுரளி
தலைவர், உத்தமம்

அனைவருக்கும் வணக்கம்

20-ஆம் நூற்றாண்டில் உருவெடுத்த கணினித்துறை கணினி அதிர, இணையம் அதிர 21ம் நூற்றாண்டில் இன்று செயற்கை நுண்ணறிவு, இயந்திர கற்றல் என பயணித்து ஆழக்கற்றலை நோக்கி அதிவேகத்தில் பயணித்துக்கொண்டு இருக்கிறது. இந்தச் சூழ்நிலையில் தமிழ்க் கணிமையும் நவீன தொழில்நுட்பத்தின் வழியே பயணிக்கச் செய்வது நமது கடமையும் கூட

கலிபோர்னியாவை தலைமையிடமாகக் கொண்டு செயல்பட்டுவரும் உத்தமம் அமைப்பு தமிழ்க் கணிமையின் வளர்ச்சிக்கு பெரும் பங்களிப்பினை செய்து வருகிறது. உலகம் முழுவதும் உள்ள தமிழ்க் கணினி தொழில்நுட்ப அறிஞர்கள் ஆர்வலர்களை கொண்டு இதுவரை 15 உலகத் தமிழ் இணைய மாநாடுகளை நடத்தி உள்ளது. இதுவரை 6 முறை தமிழகத்திலும், , 4 முறை சிங்கப்பூரிலும், இரண்டு முறை மலேசியாவிலும், அமெரிக்காவில் இரண்டு முறையும் , ஜெர்மனியில் ஒரு முறையும் ஏற்கனவே நடந்துள்ளது. இம்முறை 16வது உலகத்தமிழ் இணைய மாநாடு கனடா நாட்டில் தொராண்டோ பல்கலைக்கழகத்தில் நடைபெறுவது மகிழ்ச்சிக்கு உரியது.

"ஆழக்கற்றலும், தரவு அறிவியலும் அழகுத்தமிழில்" என்ற கருப்பொருளில் நடைபெறும் 16வது உலகத்தமிழிணைய மாநாடு மிகச்சிறப்பாக நடைபெற உள்ளது. மிகச்சிறந்த தேர்வாளர்களைக்கொண்ட ஆய்வுக்குழு சிறந்த கட்டுரைகளை தேர்ந்தெடுத்துள்ளனர். இந்த ஆய்வுக்கட்டுரைகள் நிச்சயம் தமிழ்க் கணிமைக்கு பயனளிக்கும். இந்த மாநாட்டுக்கு ஆய்வுகுழு தலைவராகவும், மாநாட்டு பொறுப்பாளராகவும் செயல்படும் பேராசிரியர் திரு.செல்வகுமார் அய்யா அவர்களுக்கும், அவர்தம் குழுவினருக்கும் , ஆய்வாளர்களுக்கும், பேராசிரியர்களுக்கும், மாநாட்டு தன்னார்வ ஊழியம் செய்துவரும் மாணவ-மாணவிகளுக்கும் எனது நெஞ்சார்ந்த நன்றி!

கணினி உலகை ஒருபுறம் Smart City, Smart Phone, Smart Village போன்ற தொழில்நுட்பங்கள் , IoT, IoE, IoV, Virtual Reality, Augmented reality, தரவு அறிவியல் என பல நுட்பங்கள் பெருகிக்கொண்டே வருகிறது.

ஆனால் தமிழ்க் கணிமை இப்போதுதான் உரை வடிவத்தில் தன்னிறைவை நோக்கி பயணித்து ஒலியிலிருந்து உரையாக்கம் கூகிளின் வழி நமக்கு அறிமுகமாகியிருக்கிறது. இன்னமும் பல துறைகளில் நாம் தமிழை கொண்டுசெல்ல பெருமுயற்சிகள் செய்வதும் அவசியமாகிறது. அதற்கான முயற்சிகளை உத்தமம் தொடர்ந்து செய்திடும்.

தமிழ்க் கணிமையை பயன்படுத்தும் ஒரு குழுவாகத்தான் நமது சமூகம் இருக்கிறதே தவிர தமிழை ஆய்வுமொழியாகவும், வணிக மொழியாக்கவும் நாம் அனைவரும் முயற்சிக்கவேண்டும்.. ஆங்கிலம் எல்லா

நாடுகளிலும் எப்படி பயணிக்கிறதோ அப்படி அதேபோல் தமிழ் சமுதாயம் தமிழை வர்த்தகம் சார்ந்த மொழியாக்கிட வேண்டும், அதன்வழியே தமிழ்க் கணிமை சார்ந்த தொழில் முனைவோர்களை உருவாக்க வேண்டும். பெரும் நிறுவனங்கள் மட்டுமே அதிக அளவில் செய்யும் பணியாக நின்றுவிடாமல் அனைத்து தொழில்முனைவோரும் தமிழ்க் கணிமையில் நிலைக்கும்படி செய்யவேண்டும்.

தமிழில் நிரலாக்கம் செய்யும் முறைகளை மிக அதிக அளவில் கொண்டுவரவேண்டும், இன்றைய சூழலில் தமிழில் நிரலாக்கம் செய்து ஒரு பயன்பாட்டு மென்பொருளையோ, குறுஞ்செயலியை உருவாக்குவதோ எளிதான செயலல்ல, இதை நாமனைவரும் ஒன்று சேர்ந்து செய்யவேண்டும். தமிழ் வளர! தமிழ்க் கணிமை வளர! நாம் அனைவரும் இணைந்து இணையம் வழித் தமிழையும் தமிழர்களையும் வளர்க்க முனைவோம்!

கூடி செய்தால் கோடி நன்மை.

வெல்லத்தமிழ் இனிதே வெல்லும்
என்றும் அன்புடன்

செல்வமுரளி

வாழ்த்து மடல்

முனைவர் வீ. சு. ராஜம்
கலிபோர்னியா, அமெரிக்கா

உத்தமம் மாநாட்டுக் கருத்தரங்குக்கு என் வாழ்த்து:

'உத்தமம்' என்னும் உலகத்தமிழ்த்தகவல் தொழில்நுட்பமன்றத்தின் சார்பில் நடக்கவிருக்கும் 16-ஆவது உலகத்தமிழ் இணைய மாநாடு இனிதே சிறப்பாக நடைபெற என் வாழ்த்து!

உடல்நிலை காரணமாக - நேரில் வந்து மாநாட்டுக்கருத்தரங்கில் பங்கேற்கும் வாய்ப்பைத் தவறவிட வேண்டியதாயிற்று.

மாநாட்டுக்கருத்தரங்கை ஒருங்கமைத்து ஒழுங்கமைத்து நேரியமுறையில் நடத்தும் பொறுப்பை ஏற்றிருக்கும் பேராசிரியர் செல்வா செல்வக்குமார் அவர்களைத் தெரியாதவர்கள் இணையத்தில் இருக்கமுடியாது.

கலை (நடனம், ஓவியம், ஒளிப்படம்...), இலக்கியம், இலக்கணம், பொறியியல், குமுகவியல், கணினியியல் எனப் பன்முகத்திறமைகொண்ட செல்வா அவர்களின் தலைமையில் இந்தக் கருத்தரங்கு செவ்வனே நடைபெறும் என்பது உறுதி!

எல்லாம் நன்றே இனிதாக நடைபெற என் நல்வாழ்த்து! என் நினைவு அங்கே இருக்கும்!

அன்பும் ஆசியும்,
ராஜம்

GPUs powering the AI revolution

Ganapathy Raman Kasi, Muthiah Annamalai
San Francisco Bay Area, California
kgrram@gmail.com, ezhillang@gmail.com

Introduction

The current hot trend in AI revolution is "deep learning" - which is a fancy way of talking about multi-layered convolutional neural networks; this field of study has heralded a new age in computing extending human capabilities by automation and intelligent machines [1].

These neural networks aren't the same as neuron networks in your brain! We are talking about artificial neural networks which reside in computers and tries to mimic the biological neural network with its synapses (connections) of axons, dendrons and their activation potentials. These thinking machines have their beginnings in post WW-II research at MIT, in the work of Seymour Papert who introduced "Perceptrons," and Norbert Weiner's "Cybernetics".

But do we know why there is sudden interest in these biologically inspired computer models ? It is due to GPUs which has accelerated all the complex computations associated with neural networks for it be practical in such a large scale. They allow these networks to operate on gigabytes (or even terabytes) of data and have significantly reduced the computation time from months to days, or days to hours, or hours to minutes usually by an order of magnitude - not possible in an earlier generation of computing. Before we jump into the details let us understand why we need deep learning and convolutional neural networks in the first place.

Scientific Innovations

Science and engineering have traditionally advanced by our ability to understand phenomena in natural world and describe them mathematically, since the times of Leonardo Da Vinci, Nicolas Copernicus, Galileo Galilei, Tycho Brahe, Johannes Kepler and Isaac Newton. However gaining models through experimentation and scientific breakthroughs piece-meal for each problem at hand is a slow process. Outside of Physics and Mathematics the scientific method is largely driven by an empirical approach.

It is in such pursuits of building models of unknown processes where observational data far exceed our human intelligence to divine an analytical model, the advent of deep learning and GPU based

multi-layered neural networks provide an ad-hoc computable model. System identification for particular classification tasks, image recognition, and speech recognition to the modern miracle of a self-driving cars are all enabled by deep learning technology. All this came about due to the seminal work of many innovators culminating in the discovery of efficient convolutional neural networks by Prof. Geoff Hinton, who trained them by hardware acceleration via GPUs.

An original pioneer in the field of AI, before the AI winter, Prof. Geoff Hinton and co-workers [2] recently showed deep learning models that beat status-quo benchmarks on classification and prediction tasks on the following speech, text or image datasets: Reuters, TIMIT, MNIST, CIFAR and ImageNet, setting off the renewed interest in the field of AI from academia and industry giants - Google, Microsoft, Baidu and Facebook alike [3].

What is a GPU ?

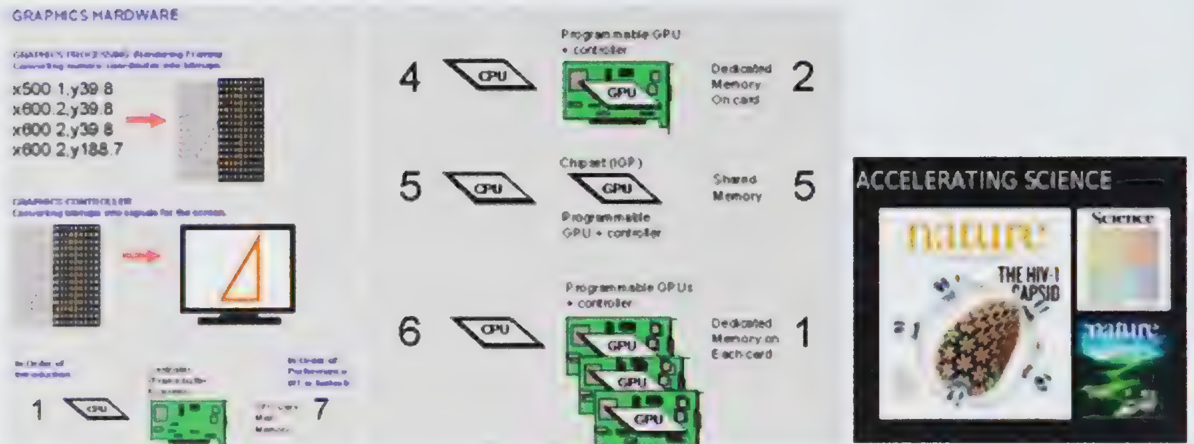
GPU stands for Graphics Processing Unit [4]. These were originally designed for graphics rendering used in video games in 1990s. They have a large number of parallel cores which are very efficient for doing simple mathematical computations like matrix multiplications. These computations are the fundamental basis for machine learning methods such as deep learning. While the improvement in CPUs over years has slowed down over the years as Moore's law has hit a bottleneck, the GPUs increase in performance has continued unabated showing tremendous improvements over the generations.



Figure. 1 (left): Deep Learning training task times as function of various GPU processors from NVidia.
Figure. 2(right): AlexNet training throughput for 20 iterations on various CPU/GPU processing platforms.

Such GPUs were originally invented for shading algorithms algorithms, are now applied in training large machine learning models using a Open CL or CUDA like frameworks (variants of C-language with description for parallel execution via threading) from the vendors.

The pioneering hardware vendors include Nvidia with their GPU series like GeForce, Tesla; AMD with its Radeon, GP GPU, Google has entered this race with its TPU (Tensor Processing Unit) and some offerings from Intel for ML training applications. Nvidia and AMD are the main players in the GPU space with Nvidia laying special emphasis on parallel computing and deep learning over the years. Nvidia just announced the new Volta generation chip based GPU V100 which is about 2.5 x faster than the previous generation chip Pascal GP100 which was announced less than 2 years ago [5].Compared to CPU, however GPUs are more than 50x faster for Deep learning. Performance of GPUs as function of various GPU families in shown in Figure. 1, and for another AlexNet data set is shown in Figure. 2.



Hardware Innovation

If the Harvard architecture and RISC architecture based CPUs have been workhorses of personal computer revolution, then the advent of high framerate video-gaming pushed the CPU based graphics rendering from CPU + Video card based rendering to CPU + GPU, to CPU + GPU + GP-GPU (general purpose GPU); some of this overview is shown in **Figure. 3a, 3b**.

Figure. 3(a,b): Evolution of GPU performance from video graphics cards and rendering from CPU; courtesy PC Magazine [4]; **Figure. 3(c):** NVIDIA Tesla GPU applications in scientific research.

Limitations

GPU's are suitable for large numerical algorithms where various data have to be moved through a computational pipeline often in parallel; this SIMD problem, like genome sequencing shown in Figure. 3c, when solved by GPU gain the maximum speedup/acceleration. However, there is a fundamental limitations of GPU acceleration due to the Amdahl's law which saturates the parallelization upto the available serial bottlenecks for a given computational task.

Software Frameworks

To build a deep learning application one may use their labeled datasets to build a learning model on any of the various frameworks [6] (both open-source or closed) provided from competing vendors in the industry as follows:

1. TensorFlow, developed by google, python API over C++ engine, low level api, good for researchers, not commercially supported; notably Google is in process of developing a TPU - an advanced version of GPU for direct use with TensorFlow.

2. Caffe 2, developed by UC Berkeley used at Facebook among other places, focused on computer vision, one of the earlier frameworks to gain significant adoption, Python API over C++ and CUDA code

3. Scikit Learn (Python based) general inference and machine-learning framework

4. Theano written in python, grand-daddy of deep learning frameworks

5. CNTK developed by Microsoft

Applications

Tamil applications for deep learning including providing or improving existing solutions to the problems of,

1. Tamil Speech Recognition
2. Tamil Character Recognition [7,8]
3. Natural Language Processing for Tamil

Hardware acceleration and availability of big-data (labeled datasets) will play key role in the success of applying deep learning techniques to these problems.

References

1. Jensen Huang, "Accelerating AI with GPUs: A New Computing Model," link: https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/

2. G. E. Hinton et-al. "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems (2012).

3. LeCun, Y., Bengio, Y. and Hinton, G. E., "Deep Learning" Nature, Vol. 521, pp 436-444. (2015),link: http://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf

4. GPU definition at PC Magazine Encyclopedia, PC Magazine, (2017) link: https://www.pcmag.com/encyclopedia/term/43886/gpu

5. Tesla GPU Application notes from NVidia, (2017) link: https://www.nvidia.com/en-us/data-center/tesla-v100/#source=pr

6. "Comparing deep learning frameworks", Deeplearning4j.org (2017), link: https://deeplearning4j.org/compare-dl4j-torch7-pylearn

7. Prashanth Vijayaraghavan, Mishra Sra, "Handwritten Tamil Recognition using a Convolutional Neural Network," NEML Poster (2015) link: http://web.media.mit.edu/~sra/tamil_cnn.pdf

8. R. Jagadeesh Kannan, S. Subramanian, "An Adaptive Approach of Tamil Character Recognition Using Deep Learning with Big Data-A Survey", Proceedings of 49th Annual Convention of Computer Society of India (vol. 1) pp 557-567 (2015), link: https://link.springer.com/chapter/10.1007/978-3-319-13728-5_63

ஆழக்கற்றல்!

(DEEP LEARNING)

முனைவர் அரிநாராயணன் சானகிராமன்

'சொன்னதைச் சொல்லுமாம் கிளிப்பிள்ளை' என்பார்கள். ஆனால், பரிணாம ஏணியில் குரங்கு மற்றும் மனிதன் உள்ளிட்ட பாலூட்டிகளுக்கு சற்றே கீழே இருக்கும் உயிருள்ள ஒரு ஜீவனான கிளிப்பிள்ளை கூட சொன்னதை மட்டும்தான் சொல்லும்! அதுவே தானாக சிந்தித்து எதையும் சொல்லுமா என்று கேட்டால் அது மிகவும் அரிது அல்லது குறைவு என்றுதான் கூறவேண்டும். ஆனால் மிகவும் சுவாரசியமாக, மனிதனால் உருவாக்கப்பட்ட, உயிரற்ற எந்திரங்களான சமகால ரோபாட்டுகள், மனிதனால் கற்றுக்கொடுக்கப்படாத பல செயல்களை மிகத் துல்லியமாகவும், பாதுகாப்பாகவும் செய்வதை நாம் கண்கூடாக காணத் தொடங்கி இருக்கிறோம்.

உதாரணமாக, செயற்கை நுண்ணறிவு கருவிகளின் வரிசையில் வரக்கூடிய மற்றும் தற்போது வெகுசன பயன்பாட்டில் உள்ள, தானியங்கி அல்லது ஸ்மார்ட் கார்கள், ஆப்பிள் நிறுவனத்தின் சிரி (ஞசைை), கூகுளின் கூகுள் நௌவ் (Googe Now), மைக்ரோசாப்ட்டின் கார்டனா (Cortana) மற்றும் அமேசான் நிறுவனத்தின் எக்கோ (நுஉாழ) உள்ளிட்ட மெய்நிகர் தனிநபர் உதவியாளர்கள் மற்றும் வீட்டுப் பாதுகாப்பு/கண்காணிப்பு கருவிகள் ஆகியவற்றைக் குறிப்பிடலாம். இக்கருவிகள் அனைத்துமே தங்களின் சொந்த அனுபவத்தில் பல விஷயங்களைக் கற்றுக்கொண்டு, தங்களின் அன்றாட செயல்பாடுகளை தாங்களே மெருகேற்றி/மேம்படுத்திக்கொள்ளக்கூடிய அசாத்திய திறன்கொண்டவை என்பது குறிப்பிடத்தக்கது!

அது சரி, மனிதர்களைப் போல செயல்படக்கூடிய இந்த புத்திக்கூர்மையும், அசாத்திய திறன்களும் இந்த செயற்கை நுண்ணறிவுக் கருவிகளுக்கு சாத்தியமானது எப்படி?

ஆழக்கற்றல் (Deep earning) எனும் அதிசயம்தான் செயற்கை நுண்ணறிவுக்கும், அதனடிப்படையில் இயங்கும் கருவிகளுக்கும் அடிநாதம்! ஆழக்கற்றல் என்றால் என்ன? மனித மூளையை பிரதிபலிக்க முயலும் பல்வேறு மென்பொருள்கள்தான் ஆழக்கற்றல் என்று அழைக்கப்படுகின்றன. நரம்பியலின் பாதிப்பில் உருவான ஆழக்கற்றல், எந்திரக் கற்றல் எனும் துறையின் ஒரு கிளைத்துறை ஆகும். ஒரு படத்தில் உள்ள பொருட்களை எல்லாம் அடையாளம் காண்பது, பேச்சு மற்றும் பல்வேறு மொழிகளை புரிந்துகொண்டு பதிலளிப்பது மற்றும் செஸ் போன்ற விளையாட்டுகளை விளையாடி உலக சாம்பியன்களையே வெகு சுலபமாக தோற்கடிப்பது ஆகியவை ஆழக் கற்றல் அடிப்படையில் இயங்கும் செயற்கை நுண்ணறிவு கருவிகளின் சாதனைகளில் 'சில' மட்டுமே என்பது குறிப்பிடத்தக்கது.

ஆழக்கற்றல் மாதிரிகளை உருவாக்குவது எப்படி?

ஆழக்கற்றலை கற்றுக்கொள்ள எந்திரக் கற்றல் குறித்த புரிதல்கள் அவசியம். எந்திரக் கற்றலில் உள்ள பாடங்கள் பெரும்பாலும், கணினிக்கான அதீத செயல்திட்டமிடல் இல்லாமலேயே சில கேள்விகளுக்கான விடைகளை கண்டறிவதாகவே இருக்கும். உதாரணமாக, 'சென்னையில் உள்ள என்னுடைய 1800 சதுர அடி

அப்பார்ட்மென்ட்டின் விலை என்ன? அல்லது 'இந்த செய்திக் கட்டுரையானது உண்மையைத்தான் கூறுகிறதா?' போன்ற கேள்விகளைக் கூறலாம். இம்மாதிரியான கேள்விகளை, எந்திரக் கற்றலில், 'X' எனும் ஏதாவது ஒரு உள்ளீட்டைக் (Input) கொடுத்தால், 'Y'எனும் அதற்கான சரியான வெளியீடு என்ன?' என்று மொழிபெயர்க்கலாம். மேலே உள்ள கேள்வியில் அப்பார்ட்மெண்ட் குறித்த தகவல்களை உள்ளீடு என்றும், கணிக்கப்பட்ட அதன் விலையை வெளியீடு என்றும் கூறலாம். எந்திரக் கற்றலில் 'மேற்பார்வை உள்ள கற்றல் (supervised learning) மற்றும் 'மேற்பார்வை இல்லாத கற்றல் (unsupervised learning) என்று இருவகை உண்டு. இக்கட்டுரையில் மேற்பார்வை உள்ள கற்றல் குறித்து விரிவாகக் காண்போம்.

மேற்பார்வை உள்ள கற்றலில், ஒரு கணினிக்கு உள்ளீடுகளுக்கான பல உதாரணங்களும், அவற்றுக்கான சரியான விடைகளும் கொடுக்கப்படும். இந்த உதாரணங்கள் 'பயிற்சித் தரவுகள் (training data)' என்றும், இந்த செயல்பாடு 'பயிற்சி (training)' என்றும் அழைக்கப்படுகின்றன. முக்கியமாக, காண்பிக்கப்படாத ஒரு உதாரணத்துக்கும் சரியான விடையைக் கண்டறிந்து கூறும் அளவுக்கு தேர்ச்சி பெறும் வரை, உதாரணங்கள் தொடர்ந்து கணினிக்கு மீண்டும் மீண்டும் பல முறை காண்பிக்கப்படும். இந்த பயிற்சியானது 'பரிசோதனை (testing)' அல்லது 'மதிப்பீடு (evaluation)' என்றும், இந்த கணினி திட்டம் 'மாதிரி' (model) என்றும் அழைக்கப்படுகின்றன.

சரி, இப்போது ஒரு சுவாரசியமான கேள்வியை உதாரணமாக எடுத்துக்கொள்வோம். ஒரு படத்தைக் கொடுத்து, 'அதில் ஒரு பூனை இருக்கிறதா இல்லையா என்று கூறு?' என்பதுதான் கேள்வி. மனிதர்களுக்கு இது மிகச் சுலபமான ஒரு கேள்வி. ஆனால், இதற்கான ஒரு கணினித் திட்டமிடலை (program) உருவாக்குவது மிக மிகக் கடினம். பாரம்பரியமான எந்திரக் கற்றல் முறைகளில், பூனையைக் குறிக்கும் பண்புகளான புசு புசுவென்று முடி, இரண்டு கண்கள், ஒரு மூக்கு கொண்ட ஒரு தலை, நான்கு கால்கள் போன்றவை காட்டப்பட்டு பயிற்சி கொடுக்கப்படும். ஆனால், கணினிக்கு காண்பிக்கப்படும் படத்தில் பூனையானது கண், மூக்கு கால்கள் எதுவுமே தெரியாமல், வால் மற்றும் உடல் மட்டுமே தெரியும்படி பின்புற கோணத்தில் இருந்தால் நாம் உருவாக்கிய கணினித் திட்டமிடல் முழுவதும் பயனற்றுப் போகும். இங்குதான் 'ஆழக்கற்றல்' முற்றிலும் மாறுபடுகிறது! எப்படி என்றால், ஆழக்கற்றலில் பண்புகள் எதையும் கொடுக்கத் தேவையில்லை. மாறாக, கேட்கப்படும் கேள்விகளுக்கான விடைகளைக் கண்டறியத் தனக்குத் தேவையான பண்புகளை, தானே கண்டறியும் திறன்கொண்ட துடிப்பான ஒரு மூளையை உருவாக்க வேண்டும்.

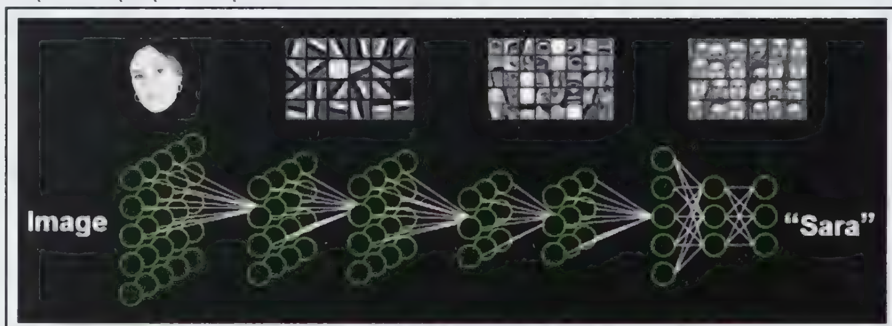
உதாரணமாக, படத்தில் (மறுபக்கத்தில்) உள்ள பூனையைக் கண்டறிய 'convolutional neural network (CNN) எனும் ஆழக்கற்றல் மாதிரி (deep learning model) வேண்டும். இந்த மாதிரியில் பல அடுக்குகள் உண்டு. முதல் அடுக்குக்கு, முழு படம் உள்ளீடாகக் கொடுக்கப்படும். உதாரணத்துக்கு, அந்தப்படமானது 200 × 200 என்ற அளவு கொண்டது என்று வைத்துக்கொள்வோம். ஆழக்கற்றல் மாதிரியின் முதல் அடுக்கானது படத்தை எடுத்து, அதனை மேல் இடது புறத்திலிருந்து கீழ் வலது புறம் வரை 10 × 10 கட்டங்களாக ஸ்கேன் செய்து அதிலுள்ள பண்புகள் தொடர்பான அனைத்து அடிப்படைத் தகவல்களையும் சேகரித்துக்கொள்ளும். பண்புகளை சேகரிக்கும் detector கருவிகளுக்கு நரம்பு (Neuron) என்று பெயர். இதில் ஆயிரக்கணக்கான நரம்புகள் இருக்கலாம். அவை ஒவ்வொன்றும் படத்தை கட்டம் கட்டமாக ஸ்கேன் செய்யும். ஒவ்வொரு நரம்பு சேகரிக்கும் தகவல்கள், 20 × 20 கட்டங்களால் ஆன பூலியன் மதிப்புகள் (boolean values) கொண்ட வலைப்பின்னல் (Grid) அல்லது சிறப்பியல்பு வரைபடங்கள் (Feature maps) என்று அழைக்கப்படுகின்றன. முதல் அடுக்கில் உள்ள அத்தனை நரம்புகளால் சேகரிக்கப்பட்ட தகவல்கள் அனைத்தும் இரண்டாம் அடுக்குக்கான உள்ளீடாகக் கொடுக்கப்படும்!

பின்னர், இரண்டாம் அடுக்கானது, முதல் அடுக்கால் அளிக்கப்பட்ட சிறப்பியல்பு வரைபடங்களை, மேல் இடது முதல் கீழ் வலது புறம் வரை கட்டம் கட்டமாக ஸ்கேன் செய்யும். ஆனால், இந்த அடுக்கில் கட்டங்கள் சற்றே பெரிய, 30 × 30 அளவு கொண்டவையாய் இருக்கும். முக்கியமாக, இரண்டாம் அடுக்கானது, பூனையின் புசு புசு வால், கண்களின் நிறம் அல்லது காதுகளின் அளவு போன்ற சிக்கலான வடிவங்கள் அல்லது பண்புளைக் கண்டறிந்து அவை தொடர்பான தகவல்களை சேகரித்துக்கொள்ளும். இதுபோல, ஒரு பூனையை முழுமையாக அடையாளம் காணத் தேவையான அனைத்துத் தகவல்களையும் சேகரிக்கும் வரை அடுக்குகளின் எண்ணிக்கையும், செயல்பாடும் தொடரும். ஆக, ஆழக்கற்றல் மாதிரியில் உள்ள ஒவ்வொரு அடுக்கும், அது சேகரிக்கும் தகவல்களும் மிக

முக்கியமானவை ஆகும்! ஆழக்கற்றலின் சிறப்பம்சம் என்னவென்றால், ஒரு பூனையை எப்படிக் கண்டறிவது என்று எதுவும் சொல்லித்தர வேண்டாம். மாறாக, ஆழக்கற்றல் மாதிரியின் கட்டமைப்பை (structure of the model) உருவாக்கி, பின்னர் அதனை பயிற்சித் தரவுகளில் இருந்து கற்றுக்கொள்ளும்படி தயார் செய்துவிட்டால் போதும். கேள்விகளுக்கான விடைகளைத் தன்னுடைய நரம்புகளைக் கொண்டு தானே கண்டுபிடித்துக் கொள்ளும்.

என்ன நண்பர்களே, ஆழக்கற்றல் பாடங்களை படிக்க இப்போது ஒரு பேரார்வம் வந்துவிட்டது தானே? கண்டிப்பாக படிக்கலாம்! ஆழக்கற்றல் பாடத்திட்டங்கள் ஆன்லைன் மற்றும் கல்வி நிறுவனங்களில் தற்போது பரவலாகக் கற்றுத்தரப்படுகின்றன.

முகத்தைக் கண்டறியும் மற்றொரு உதாரணம்:



A figure from nvidia dev blog.

எண்ணிம நூலகம் / ஆவணகம் தமிழகத்தின்தொல்லியல் சின்னங்கள்

மு. உதயசங்கர்

எண்ணிம நூலகம் (digital library) என்பது எண்ணிம அல்லது மின்னியல் முறையில் நூல்கள், படங்கள், ஆவணங்கள், தகவல் தொகுப்புகளைச் சேகரித்து மின்னியல் முறையில் பாதுகாத்து வைக்கக் கூடிய முறையே எண்ணிம நூலகம் ஆகும். மின் நூலகம் (electronic library), மெய்நிகர் நூலகம் (virtual library) போன்ற பெயர்களாலும் அழைக்கப்படுகிறது. இந்த நூலகத்தில் உள்ள எண்ணிம உள்ளடக்கங்களை (Digital content) கணினி இணையம் வழி அணுகிப் பயன் பெறலாம். நூல்கள், ஆவணங்கள், ஒலி, ஒளி தகவல்களை மிக விரிவாக எண்ணிம உள்ளடக்கங்களாகச் சேகரித்து, மேலாண்மை செய்து, பாதுகாத்து பயனர்களுக்கு அத்தொகுப்புகளை தேவைப்படும் போது பயன்படுத்த அல்லது தரவிறக்கம் செய்வதற்கேற்ப வடிமைக்கப்படுகிறது.

அச்சில் இல்லாத நூல்கள், வரலாற்றுச் சிறப்பு மிக்க பழமை வாய்ந்த நூல்கள், ஆவணங்கள், குறிப்பேடுகள், செய்தித்தாள்கள், படங்கள் என எளிதில் அழியக்கூடியவைகளை படங்கள், ஒலி, ஒளித் தகவல்களோடு மேலும் கிடைத்தற்கரிய தகவல்களை செம்மைப்படுத்தி சேமித்து வைப்பதால் அவற்றை அழிவிலிருந்து காக்க முடியும்.

தமிழகத்தின் தொல்லியல் சின்னங்கள் எண்ணிம நூலகம் / ஆவணகம் தேவையும் பயன்பாடும்:

மனித இனம் தோன்றி அவற்றின் படிப்படியான வளர்ச்சியில் தனியாகவும் குழுவாகவும் வாழத் தொடங்கிய காலத்தில் உணவுத் தேடலுக்காகவும், பிற தேவைகளுக்காகவும் அலைந்து திரிந்த மனிதர்கள் ஒருவருக்கொருவர் தமக்கு ஏற்பட்ட அனுபவங்களைத் தம் குழுவினருடன் பகிர்ந்து கொள்ள குகைகளிலும் பாறைகளிலும் சித்திரமாக வரைந்ததே மனிதன் தன் இருப்பை ஆவணப்படுத்தத் தொடங்கியதன் முதல் படியாகும்.

ஆரம்ப காலத்தில் இலக்கியங்கள் வாய்மொழியாக இருந்தது. பின்னர் தனக்குத் தெரிந்தவற்றை அழியாமல் பதிவுசெய்து பாதுகாக்க வேண்டும் என்ற எண்ணம் தோன்றியதோ அப்போது எழுத்து வடிவம் தோன்றியது எனலாம். எண்ணங்கள் முதலில் கோடுகளாக, குறியீடுகளாக, படங்களாக இருந்து பின் ஒவ்வொரு ஒலிக்கும் ஒரு வடிவாக உருப்பெற்று பின் படிப்படியாக வளர்ந்து தற்போதைய எழுத்து வடிவமாக மாறின என்பதற்குப் பாறைகள், கல்வெட்டுகள், செப்பேடுகள், சுவடிகள் போன்றவை சான்றுகளாக உள்ளன.

களிமண் பலகை, பாறை, கல், உலோகத்தகடு, துணி, இலை, பனைஓலை, மரப்பலகை, மிருகதோல், போன்றவை எழுதப்படும் பொருட்களாகப் பயன்பட்டன. இலை, மரப் பலகை, களிமண் பலகை போன்றவை விரைவில் அழியக் கூடியவை மேலும் பெரிய நூல்களை எழுதிக் கையாளுவது கடினம். தோல், துணி, உலோகத்தகடு போன்றவை பொருட்செலவினை உண்டாக்கும். அவற்றில் விரைவாக எழுதவும் முடியாது. கருங்கல் போன்ற பிறபொருள்களைப்

பிற இடங்களுக்கு எடுத்துச் செல்வது கடினம். ஆனால் ஓலைச் சுவடியோ சரியான முறையில் கையாண்டால் இரண்டு மூன்று நூற்றாண்டுகள் வரை அழியாதவாறு பாதுகாக்க முடியும். அதிக செலவு இல்லாது, மிகுதியாகவும், எளிமையாகவும் கிடைக்கக் கூடியது. மிகப் பெரிய அளவுடைய நூல்களையும் ஒரு கட்டில் அடக்கலாம்.

ஆவணப்படுத்துவதின் அவசியம் புரிந்ததால் தான் நீண்ட காலம் அழியாதிருக்க வேண்டுமென்றும் தேவை, அவசியத்தின் பொருட்டு அதற்குத் தகுந்தவாறு எழுதப் பயன்படும் பொருட்களை தேர்ந்து எடுத்தனர் நம் முன்னோர்கள்.

தகவல்களைக் கடத்த எழுதப்படும் பொருட்களுக்கு இம்முக்கியத்துவம் தந்தவர்கள் ,மக்களின் வாழ்வியல் சார்ந்த கூறுகளையும் சிற்பங்கள், உலோக உருவங்கள், மரத்திலான பொருட்கள், கம்பளங்கள், ஓவியங்கள், அலங்காரப் பொருட்கள், கண்ணாடி, பீங்கான், படைப்புகள் போன்றவற்றை உருவாக்கி விலை மதிப்பற்ற கலை பொக்கிஷங்களாகத் தந்துள்ளனர். பல்வேறு சமயங்களும், வழிபாட்டு இடங்களும் நேரடியாகவும், மறைமுகமாகவும் கலைகள் உருவாவதற்கு உதவியாக இருந்து வந்துள்ளன. அவ்வாறு கிடைக்கப் பெற்ற பொருட்கள் மூலம் அன்றைய காலகட்டத்தின் வரலாற்றை ஆய்வாளர்கள் கணிக்கிறார்கள்.

ஒவ்வொரு பொருளும் இயற்புச் சிதைவு உட்பட தட்பவெப்ப நிலை, ஒளி, பூச்சிகள், காளான்கள், பூஞ்சைகள் போன்ற பல்வேறு காரணிகளால் சிதைவிற்கு உள்ளாகிறது. தொன்மையைப் பற்றிப் புரியாதவர்கள் அதை சிதைத்து வருகின்றனர். அழிவின் விளிம்பில் இருக்கும் நம் முன்னோர் வாழ்வியல் சின்னங்களை அடுத்த தலைமுறைக்கு அப்படியே தர முடியாவிட்டாலும் இயன்றவரை இன்றைய கணினி யுகத்தில் எண்ணிம தகவலாக ஆவணப்படுத்த வேண்டிய கட்டாயத்தில் உள்ளோம்.

தமிழகத்தின் மிகச் சிறந்த படைப்பாக்கங்களாகவும், பாரம்பரிய சின்னங்களாகவும் கருதப்படும் கோவில்கள், கல்வெட்டுகள், சிற்பங்கள், குளங்கள், வழிபாட்டு இடங்கள், பழங்காலக் கட்டிடங்கள், ஓலைச்சுவடிகள், மக்கள் குடியிருப்பு பகுதிகள், ஈமசின்னங்கள் உள்ள இடங்கள், நடுகல், கல்திட்டை, கற்குவை, அரண்மனைகள், கோட்டைகள், போர்க்கருவிகள், நாணயங்கள், மக்களின் பயன்பாட்டிற்கான அன்றாட பொருட்களின் மூலம் அவை அமையப் பெற்ற நிலம், மலைகள், ஆறு, ஏரி, கடல் மற்றும் அருவிகள் போன்ற அக்கால மக்களின் வாழ்வியலோடு இணைந்த நிலத்தோற்றங்களையும் கருத்தில் கொண்டு இடங்காட்டி (GPS Locat on) தொழில்நுட்ப உதவியுடன் தனியாக ஒரு தரவுத்தளத்தினை மன்னர், மன்னரின் பரம்பரை (aynasty), அரசாண்ட காலம் இவற்றோடு தொடர்புபடுத்தி சரியான தரவுகளோடு அல்லது மாற்று கருத்து இருப்பின் அவைகளையும் சேர்த்து உருவாக்கி ஆவணப்படுத்துவது.

சேகரிக்கப்பட்ட தகவல்களை அவை சார்ந்த புற விடயங்களுடன் வரலாறு, கலை, இலக்கியம், புவியியல், நிலவியல், இயற்பியல், தகவல் அறிவியல், வேதியியல், புள்ளியியல், சூழலியல், விலங்கியல், தாவரவியல் போன்ற துறைகளையும் தளமாக படுத்திக் கொள்வதன் மூலம்:

- தமிழகத்தின் வரலாற்றை காலவரிசை அட்டவணைப்படுத்த முடியும்.
- கல்வெட்டுகளில், ஓலைச்சுவடிகளில் உள்ள தகவல்களை தேடுபொறி (Search Engine) கொண்டு எளிதாக தகவல்களை பெறுவது.
- மன்னர் / மன்னர் பரம்பரையினர் வாழ்ந்த இடங்கள், போர் நடந்த இடங்கள், வென்ற பகுதிகள், போர்க்கால நினைவுச் சின்னங்கள், அவர்கள் எழுப்பிய / மறு சீரமைப்பு செய்த கோவில்கள், கல்வெட்டுப் பாடல்களை ஆராய்ந்து அவர்கள் ஆண்ட பகுதிகளின் எல்லைகள் எது என வரைபடம் (Mapping) பெறுவது.
- மன்னர் / மன்னர் பரம்பரையின் காலத்து பண்பாட்டு அம்சங்கள், சமூகநிலை போன்ற தகவல்களையும் பெற உதவும்.
- பண்பாட்டுப் படிநிலை வளர்ச்சியைக் கால வரிசைப்படுத்தல்.
- வருங்கால ஆராய்ச்சியாளர்களுக்கு எல்லாவகையான தரவுகளும் ஒரே இடத்தில் கிடைக்க வழிவகை

செய்யும்.

- தமிழக வரலாற்றின் அரசியல், சமூக, பொருளாதார நிலையினை முதன்மைச் சான்றுகளான தொல்பொருட்கள், கல்வெட்டுகள் கொண்டு அதன் தொன்மையை வெளிக்கொணர்தல்.
- வரலாற்றுக்கு முந்திய காலமான பழைய கற்காலம், இடைக்கற்காலம், புதிய கற்காலம் போன்றவற்றின் தரவுகளோடு அதனதன் தொன்மையை வெளிக்கொணர்தல்.
- வரலாற்று ஆய்வாளர்கள், தொல்லியல் துறையினர், புவியியல் துறையினருக்குமான பாலமாக உருவாக்குதல்.

வரலாற்றுக்கு முந்திய மற்றும் வரலாற்றுக் கால மனிதப் பண்பாட்டின் தோற்றத்தையும், வளர்ச்சியையும் ஆவணப்படுத்தல், அவற்றை விளக்குதல், அதனை தற்போது இருக்கும் தொழிற்நுட்பத்தை கொண்டு சரியான முறையில் அட்டவணைப் படுத்துவதன் மூலம் வரலாற்றைப் எளிதில் புரிந்து கொள்ளுவதற்கு உதவும்.

ஆய்வாளர்களுக்கு மட்டுமல்லாமல் பொது பயனர்களும் அவர்களுக்கு தேவையான தனித் தகவல்களை நேரடியாக ஒப்பீடு செய்து கொள்ளலாம்.

www.udhayam.in என்ற பெயரில் தொடங்கப்பட்டுள்ள இந்த இணையதளத்தில், பழங்காலக் கோவில்கள், குகைகள், மலைகள், சின்னங்கள் போன்றவை அதன் தரவுகளோடு அதற்குண்டான புகைப்படங்கள், மற்றும் ஒளிப்படங்களுடன் விளக்கமாகவும் (Image With Description) நகரமயமாக்கலில் காணாமல் போன கோலங்கள் வரைவது எப்படி என்ற அசைப்படங்களும் (Animation) ஒளிப்படங்களாக்கப் பதிவேற்றப்பட்டுள்ளது. இத்தளத்தில் தொன்மையான கலைகளை ஆவணப்படுத்தவும் மற்றும் தமிழின் பழங்கால எழுத்து வடிவங்களை எழுத்துருக்களாக மாற்றும் முயற்சியிலும் உள்ளேன்.



உத்தமம் பற்றி

அமெரிக்க நாட்டில் பதிவு செய்யப்பட்ட இந்நிறுவனம், "உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம் (உத்தமம்)" என்னும் பெயரில் இயங்கி வருகிறது. இந்த நிறுவனத்தின் முக்கிய குறிக்கோள்களில் ஒன்று தமிழ்த் தொழில் முனைவோர், தமிழ்க் கணினியாளர்கள், நிரலர்கள், பேராசிரியர்கள், ஆசிரியர்கள் மற்றும் மாணவர்கள் ஆகியோரை ஒருங்கிணைக்க முனைவதாகும். இதன் உறுப்பினர்கள் இந்தியா, இலங்கை, சிங்கப்பூர், மலேசியா, வட அமெரிக்கா மற்றும் ஐரோப்பிய நாடுகளில் உள்ளனர். இந்நிறுவனத்தின் ஒரு முக்கிய நடவடிக்கை ஆண்டுதோறும் தமிழ் இணைய மாநாட்டை உலகெங்கிலும் பல நாடுகளில் தொடர்ந்து நடத்திவருவதாகும்.



ஆழக்கற்றலும் கணினிமொழியியலும்

(Deep learning and computational linguistics)

இராசேந்திரன் சங்கரவேலாயுதன்
அமிர்தாப் பல்கலைக்கழகம், கோயம்புத்தூர்
rajushush@gmail.com

1. வாழ்த்து

தமிழிணையம் தனது ஆறாவது மகாநாட்டை 2003–ஆம் ஆண்டு சென்னையில் நடத்தியபோதுதான் தமிழிணையத்தோடு இணைய எனக்கு வாய்பு கிடைத்தது. அம்மாநாட்டில் தமிழ்ச் சொல்வலை (Tamil wordNet) குறித்து முதல்முறையாக உலக அரங்கில் பேச எனக்கு ஒரு வாய்ப்புக்கிடைத்தது. அதன் பின்னர்தான் தமிழ் இணையத்தின் செயல்பாடுகள் குறித்து அறிந்துகொண்டேன். 2010–இல் கோயம்பத்தூரில் நடைபெற்ற தமிழிணைய மாநாடு அறிவு உருப்படுத்திற்கு கருத்துருசார் அகராதி என்ற கட்டுரையைச் சமர்ப்பிக்கச் சந்தர்ப்பமளித்தது. 2013–கோலாலம்பூரில் மலேயாப் பல்கலைகழத்தில் நடைபெற்ற தமிழிணைய மாநாடு கணிசார் தமிழ்ச் சொல் மயக்கம் நீக்கம் என்ற கட்டுரையைச் சமர்ப்பிக்க வாய்ப்பளித்தது. 2017–இல் காந்திகிராம கிராமியப் பல்கலைக்கழகத்தில் நடை பெற்ற இணையமாநாடு தமிழ்ச் சொல்வலையிலிருந்து காட்சி மூலப்பொருண்மையியல் சொற்களஞ்சியம் என்ற தலைப்பில் பேச ஒரு வாய்ப்பளித்தது. தமிழ்க் கணினிமொழியியல் குறித்து உலகளாவிய ஒரு அரங்கத்தில் பேசவும் கட்டுரைகளை வழங்கவும் எனக்கு வாய்ப்பளித்தது தமிழிணையம் தான். தமிழிணைய மாநாடுகள் தான் எனது கருத்துக்களையும் கட்டுரைகளையும் உலகளவில் கொண்டுசெல்ல உதவிசெய்து வருகின்றது என்பதை நான் இங்கு பதிவு செய்துகொள்கின்றேன். இத்தகைய இணையமாநாடுகள் சிறப்புற நடைபெற எனது வாழ்த்துகள்.

2017–தொரந்தோவில் நடைபெறவிருக்கும் மாநாட்டில் எனக்குத் தரப்பட்டுள்ள வாய்ப்பை அம்மாநாட்டின் கருத்துமுழக்கமான ஆழக்கற்றல் என்பது பற்றிய எனது கருத்துக்களை முன்வைக்கலாம் என எண்ணுகின்றேன்

2. தொடக்கவுரை

மொழியியல் என்பது மனித நாகரீகத்தின் மிக பழமையான துறையாகும். நாம் பல்லாயிரக்கணக்கான ஆண்டுகளாக மனிதமொழியின் வேறுபட்ட நோக்குகளை ஆய்ந்து வருகின்றோம். பல நேரங்களில் மனித மொழி அறிவு, புலனறிவு மற்றும் கருத்துப் பரிமாற்றம் இவை குறித்து ஆய்வதற்கு முக்கியத்துவம் தரப்பட்டது. பல நூற்றாண்டுகளாக இத்துறை மனித அறிவின் பிற கிளைகளுடன் அதன் கருத்துருசார் தொடர்பை நிறுவிப் படிப்படியாக அதன் முழு வடிவில் வளர்ந்துள்ளது. இப்போது, இன்றைய நூற்றாண்டின் தொடக்கத்தில் இது எவ்வாறு மொழியின் வேறுபட்ட நோக்குகளைப் பற்றிய கோட்பாடுகள் சீரான மனித மொழி வெளிப்பாடுகளில் உருப்படுத்தம் செய்யப்பட்டுள்ள மொழியின் உண்மையான பயன்பாட்டிற்குச் சான்றாக மெய்படுத்தம் செய்யப்பட்டுள்ளது என்பதை ஆய்வதில் தன் கவனத்தைத் திருப்பியுள்ளது. மொழி ஆய்வின் இப்புதிய திசை மரபு

மொழியியலுக்குக் கூடுதல் பரிமாணத்தைத் தந்துள்ளது. மொழியியல் ஆய்வுக்குப் புதிய கருவிகளின் குழுமத்தையும் மொழியின் உண்மையான பயன்பாட்டின் எடுத்துக்காட்டுகளைச் சேகரிக்கும் தொழில் நுட்பங்களையும் நல்கியக் கணிப்பொறித் தொழில் நுட்பத்தின் அறிமுகத்தால் இது சாத்தியமானது. இப்புதிய அணுகுமுறையின் அறிமுகம் மொழியிலுக்கும் மனிதவினத்திற்கும் இரண்டு வழிகளில் நன்மைபுரிந்துள்ளது. முதலாவது இது மொழி மற்றும் மொழிப்பயன்பாடு பற்றிய பழைய கோட்பாடுகள் தொடர்ந்து பயன்படுத்த உகந்ததா என்பதைத் தெரிந்துகொள்ள உதவுகின்றது. இரண்டாவது இது சீரான மனிதச் செயல்பாடுகளில் மொழிச் சான்றின் மற்றும் தகவலின் நேரடியான பயன்பாட்டிற்கு அனுகூலங்களைத் தருகின்றது. இது சரியான திசை, மாறுபாடு மற்றும் பயன்பாடு என்பனவற்றின் குறைவால் சிலகாலம் பாதிக்கபட்டுள்ள மொழியலின் மறுமலர்ச்சிக்கும் நிலைநிற்புக்கும் உதவுகின்றது. சில காலகமாகக் கணிப்பொறித் தொழில்நுட்பத்தின் கண்டுபிடிப்பு மற்றும் முன்னேற்றம் மொழியிலுக்கு ஒரு புதிய பரிமாணத்தைத் தந்துள்ளது.

3. பண்புசார் மொழியியலும் அளவுசார் மொழியியலும்

மொழியியல் அதன் தொடக்கக் காலகட்டத்தில் மொழித் தரவைச் சேகரித்து மொழியியல் கோட்பாடு மற்றும் விதிகள் அடிப்படையில் அதை ஆய்ந்து முடிவுகளை எய்துவதை வழக்கமாகக் கொண்டிருந்தது. சாம்ஸ்கி மொழியின் ஒருபகுதியை மட்டும் உருப்படுத்தம் செய்யும் குறைவான தரவு அடிப்படையிலான அணுகுமுறை தவறானது என்றும் அது மொழிகுறித்த முழுவதுமான முடிவைத் தராது மற்றும் குறைபாடு உடையது என்ற வாதத்தை முன்வைத்து சரியான முடிவுகளை எய்த மொழிபேசுபரின் அக அறிவைப் பார்க்கவேண்டும் என்ற கோட்பாட்டைப் பரிந்துரைத்தார். கணினியின் வரவால் பெருந்தரவுகளை/விரிதரவுகளை அல்லது தரவுத்தொகுதிகளை சேகரித்து எளிதில் மொழியாய்வு செய்ய இயலும் என்ற சாத்தியம் காரணமாகவும் சாம்ஸ்கியின் கோட்பாடு கோட்பாடு அடிப்படையில் சரியாக இருந்தாலும் மொழி பேசுபவரின் அறிவை மூளையைப் பார்த்து அறியமுடியாது அவர் வெளிப்படுத்தும் மொழித்தரவுதான் கைகொடுக்கும் என்ற காரணத்தால் மரபுமொழியியல் ஆய்வு தரவுத்தொகுதி மொழியியல் ஆய்வாக (Corpus linguistics) மறு உருப்பெற்றது. தரவுத்தொகுதி மொழியியல் ஆய்வு முந்தைய தரவு அடிப்படையிலான ஆய்வில் உள்ள குறைபாடுகளை நிவர்த்தி செய்யும் என்ற வாதம் முன்வைக்கப்பட்டது. மொழியியல் ஆய்வு பண்புசார் மொழியியலிலிருந்து (qualitative linguistics) அளவுசார் மொழியியியலாக (quantitative linguistics) மாற்றம் பெற்றது. அளவுசார் மொழியியல் முடிவுகள் புள்ளியல் அடிப்படையில் அமையும்.

4. கணினிமொழியியலும் இயற்கைமொழி ஆய்வும்

மொழியைப் புலனறிவுடன் தொடர்புடைய மனிதக் கருத்துப்பரிமாற்றத்தின் கருவியாகப் பார்ப்பதை நோக்கமாகக் கொண்ட செயற்கை அறிவுநுட்பத்தின் (artificial intelligence) பகுதியாகக் கணினி மொழியியல் (computational linguistics) என்ற பதிய பாடம் உருவாகியது. விரிதரவு மொழியியல் கணினிமொழியியலின் முக்கியமான கிளையாகப் புள்ளியியலில் பயன்படுத்தப்படும் நெறிகள் மற்றும் உத்திகளைப் பின்பற்றி மிக ஒழுங்கான வழியில் சேகரிக்கப்பட்டுள்ள மொழிப் பயன்பாட்டின் அனுபவவாதச் சான்றின் பெரும் அளவைத் தரும் செயல்பாட்டில் முக்கியப் பங்களிப்பு செய்கின்றது. மேலும் இது விரிதரவை ஆயவும் மனித மொழியைப் புரிந்துகொள்ளவும் மனித அறிவியலின் பல களங்களில் பயன்படுத்தவும் கணினி மொழியியலிலும் செயற்கை அறிவுநுட்பத்திலும் தேவையானதாகக் கருதப்படும் தகவல்களைப் பெற மிக நுட்பமான உபாயங்களின் குழுமத்தைத் தருகின்றது. எவ்வாறு மக்கள் மொழியை ஒரு கருவியாகப் பயன்படுத்தி தங்களுக்குள் கருத்துப்பரிமாற்றம் செய்கின்றனர் என்பதைப் புரிந்துகொள்ள வலுவான புலனறிவு மற்றும் மொழியியல் சார்ந்த நோக்கம் இருக்கின்றது. மனிதவினத்துடன் திறமையாக மொழி ஊடாட்டங்கள் செய்யவியலும் அறிவுநுட்பமுள்ள கணிப்பொறி ஒழுங்குமுறைகளை உருவாக்கத் தொழிநுட்டபம் அடிப்படையிலான நோக்கம் இருக்கின்றது. இந்த நோக்கங்களுடன் கணிப்பொறி அறிவியலார்களும் மொழியலார்களும் ஒன்றாக இணைந்து இயந்திர மொழிபெயர்ப்புக்கும் தகவல் பிரித்தெடுப்புக்கும் மொழி புரிதலுக்கும் மொழி உருவாக்கத்திற்கும் ஒழுங்குமுறைகளை உருவாக்கியுள்ளனர். இச்செயல்பாடுகளுக்கு அவை சீரான மற்றும் அரிதான பண்புக்கூறுகளைக் கொண்ட இயற்கை மொழியை அனுபவவாத அடிப்படையில் புரிந்துகொள்ள வேண்டியது தேவையாகும். இங்கு விரிதரவு தவிர்க்க முடியாததாக மாறுகின்றது; ஏனென்றால் விரிதரவு அனுபவவாதத்

தரவுத்தொகுதிகளின் சேகரிப்புக்குள் அறியவியலும் மொழியின் பெரும்பாலான பண்புக்கூறுகளை வெளிக்கொணர திறன்களைக் கொண்டுள்ளது. தற்போது பல எண்ணிக்கையிலான மக்கள் கணிப்பொறியில் பல்வேறு வகைப்பட்ட மொழியியல் தகல்களை நடைமுறைப்படுத்துவதில் ஈடுபட்டுள்ளனர் ஏனென்றால் கணினி மொழியியலின் மற்றும் இயற்கை மொழியாய்வின் (natural language processing) நோக்கங்கள் மொழியின் பண்புக்கூறுகளை முடிந்த அளவு கணினி அடிப்படையில் பண்பாக்கம் செய்யவேண்டும் என்பதாகும். மேலும் இயற்கை மொழியியல் ஆய்வுக் களத்தில் செய்யப்பட்டுள்ள ஆராய்ச்சிகள் ஒரு இயற்கை மொழியின் மொழியியல் பகுப்பாய்வில் வெளிப்படுத்தப்படும் உள்ளறிவிலிருந்து நன்மையடைந்துள்ளன. எடுத்துக்காட்டாக, ஒரு கணினியால் தானாக ஒரு எளிய வாக்கியத்தைப் பொருள்கோள் செய்ய ஒழுங்குமுறையை உருவாக்குபவர் இம்மாதிரி வாக்கியங்களின் முந்தைய மொழியியல் பகுப்பாய்வு பயனுள்ளதாகவும் அறிவுபுகட்டுவதாகவும் இருப்பதைக் காண்பார். எனவே ஒரு விரிதரவுக்குள் சேகரிக்கப்பட்டுள்ள மொழி அலகுகளின் வர்ணனைகளும் தன்மைகளும் மொழித் தொழில் நுட்பத்திற்கு (language technology) மிக முக்கியமானவைகளாகும். விரிதரவிலிருந்து எடுக்கப்படும் புதிய தகவல்கள் கணினி மொழியியலுக்குப் பயன்படுவதுடன் முக்கிய மொழியியல் களத்திற்குப் பயன்படும் மொழியின் வருணனை மற்றும் புரிந்துகொள்ளல் பற்றிய மதிக்கத்தக்க உள்ளறிவுகளையும் தருகின்றது. மொழியில் மொழி ஊடாட்டம், கருத்துப்பரிமாற்றம் மற்றும் புரிந்துகொள்ளல் என்ற பல்வேறு களங்களில் வெளிப்படுத்தப்பட்டுள்ள ஒரு இயற்கை மொழியின் பயன்பாட்டின் எல்லா திசைகளையும் உள்ளடக்கும் விரிந்த பரப்பைக் கொண்ட பல்பரிமணக் களமாகும். மொழி ஆய்விலும் பயன்பாட்டிலும் விரிதரவின் அறிமுகம் மைய மொழியிலுக்குப் புதிய பரிமாணங்களை உட்படுத்தியுள்ளது. கொள்கை அடிப்படையில் விரிதரவு மொழியியல் (corpus linguistics) இயந்திரத்தால் படிக்கவியலும் உரைகளின் மிகப்பெரிய சேகரிப்புகளின் வழி மொழியியல் நடப்புகளை ஆய்வதை நோக்கமாகக் கொண்ட அணுகுமுறையாகும். இவ்வணுகுமுறை பல ஆய்வுக்களங்களில் பயன்படுத்தப்படுகின்றது: ஒரு மொழியின் வருணனை ஆய்விலிருந்து மொழித் தொழில் நுட்பம் மற்றும் கல்வி வரையிலான எல்லா ஆய்வுகளிலும் பயன்படுத்தப்படுகின்றது. விரிதரவு மொழியியலின் உள்ளூறையும் கருத்துரு விரிவாகப் பலவற்றை உள்ளடக்கியதாகும். இது நம்பகமான பேச்சு மற்றும்/அல்லது எழுத்து உரை மாதிரிகளின் எந்தக் கணிசமான பகுதியின் முழுமையான ஆய்வையும் குறிப்பிடும். இது மொழியின் ஒருகால மற்றும் இருகால நோக்குகளை வருவரைவு செய்யும் இலக்கிய உரைகள் மற்றும் செய்தித்தாள் கட்டுரைகள் இவற்றின் சேகரிப்புகளை உள்ளடக்கிய தற்கால மொழியின் பெரிய அளவிலான இயந்திரத்தால் படிக்கவியலும் விரிதரவுகளை உட்படுத்தும். மின் விரிதரவு (electronic corpus) என்பது புதிய ஒன்றாகும். இக்களம் அரை நூண்றாண்டுக்கும் மேற்பட்ட வரலாறு உடையது. எது ஒரு விரிதரவாகக் கருதப்பட வேண்டும், அது எவ்வாறு திட்டமிடப்படவேண்டும், எவ்வாறு உருவாக்கப்படவேண்டும், வகைப்படுத்தப்படவேண்டும், ஆயப்படவேண்டும் மற்றும் பயன்படுத்தப்படவேண்டும் என்பதில் பொதுவான கருத்து இல்லை.

5. பெருந்தரவிலிருந்து மொழியியல் சார் முடிவுகளைப் பெறல்

கணினிமொழியியல் அடிப்படையில் முடிவுகளைப்பெற குறைந்தது மூன்று அல்லது நான்கு முறைகள் வழக்கில் உள்ளன: கண்காணிக்கப்பட்ட முறைகள், கண்காணிக்கப்படாத முறைகள், கலப்பு முறைகள், அறிவு அடிப்படியிலான முறைகள். கண்காணிக்கப்பட்ட முறைகள் மொழியியல் செய்திகளை வேண்டும் கண்காணிக்கப்படாத முறை எவ்வித மொழியியல் செய்திகளையும் வேண்டாது பனுவல் அல்லது தரவிலிருந்து கணிப்பொறி தானாகவே பெற இயலும் தகவல்கள் அடிப்படையில் முடிவுகளை எய்தும். கலப்புமுறை கண்காணிக்கப்பட்ட முறையையும் கண்காணிக்கப்படாத முறையையும் உட்படுத்தும். அறிவு அடிப்படையிலான முறை அறிவு அடிப்படையிலான மூலங்களை வேண்டும்.

கண்காணிக்கப்பட்ட முறைகள் இயந்திரம்கற்றல் முறையில் அமையலாம். அதாவது தரப்பட்ட பனுவலுக்கு மொழியியல் தகவல்களை மனிதமுயற்சியால் தருவது கடினமானதும் காலவிரையமானதும் ஆகும். இதற்காகப் பனுவலின் ஒரு சிறிய பகுதிக்கு மொழியியல் தகவலைத் தந்து அதற்குப் பின்னணியாக அமையும் விதிகளை மறைமுகமாக இயந்திரத்தைக் கொண்டு கற்கச்செய்து பின்னர் தானாகவே அது மொழியியல் தகவல்களை மீதமிருக்கும் மொத்த பனுவலுக்கும் தருவது இயந்திரம் கற்றல் முறையாகும். எடுத்துக்காட்டாகப் பனுவலின் ஒரு சிறிய பகுதியை மனித முயற்சியால் சொல்வகைபாட்டிற்கு அடையாளப்படுத்தித் தந்தால் கணினி அதன் கற்றல்

திறன்கொண்டு மீதி இருக்கும் பனுவலைத் தானியக்கமாகச் சொல்வகைப்பாட்டிற்கு அடையாளப்படுத்திக் கொள்ளும்.

6. இயந்திரம் கற்றலும் ஆழக்கற்றலும்

ஆழக்கற்றல் ஆழமான கட்டமைக்கப்பட்ட கற்றல் (deep structured learning) அல்லது படிநிலைக் கற்றல் (hierarchical learning) எனவும் அறியப்படுகிறது. ஆழக்கற்றல் என்பது செயற்கை நரம்புப் பிணையங்களை (Artificial Neural Networks (ANNs)) ஒன்றுக்கு மேற்பட்ட மறைக்கப்பட்ட அடுக்குகளைக் கொண்டிருக்கும் கற்றல் பணிகளுக்குப் பயன்படுத்துவதாகும். குறிப்பிட்ட பணி நெறிமுறைகளுக்கு மாறாகக் கற்றல் தரவு உருப்படுத்தங்கள் அடிப்படையிலான இயந்திரக் கற்றல் முறைகளின் பரந்த குடும்பத்தின் ஒரு பகுதி ஆழக்கற்றலாகும். கற்றல் கண்காணிக்கப்பட்டதாகவோ பகுதியளவு கண்காணிக்கப்பட்டதாகவோ கண்காணிக்கப்படாததாகவோ இருக்கலாம்.

சில உருப்படுத்தங்கள் பல்வேறு தூண்டுதல்களுக்கும் மூளையில் தொடர்புடைய நரம்பியல் பதில்வினைகளுக்கும் இடையில் ஒரு உறவை வரையறுக்க முயற்சிக்கும் நரம்பியல் குறியீட்டு போன்ற உயிரியல் நரம்பு மண்டலத்தில் உள்ள தகவல் செயலாக்க மற்றும் தகவல்தொடர்பு அமைப்பொழுங்கின் விளக்கத்தை அடிப்படையாகக் கொண்டவை. பெரிய அளவிலான, அடையாளப்படுத்தப்படாத தரவுத் தொகுப்பிலிருந்து இந்த உருப்படுத்தங்களைக் கற்கத் திறமையான அமைப்புகளை உருவாக்க ஆராய்ச்சி முயற்சிக்கிறது.

ஆழமான நரம்பியல் பிணையங்கள், ஆழமான நம்பிக்கை பிணையங்கள் மற்றும் மீளநிகழும் நரம்பியல் பிணையங்கள் போன்ற ஆழக்கற்றல் கட்டமைப்புகள் கணினிப் பார்வை, பேச்சு அறிதல், இயற்கை மொழி ஆய்வு, கேட்பொலி அறிதல், சமூகப் பிணையங்களை வடிகட்டுதல், இயந்திர மொழிபெயர்ப்பு மற்றும் உயிர் தகவலியல் போன்ற களங்களுக்குப் பயன்படுத்தப்படுகின்றது இவற்றில் அவை மனித வல்லுநர்களுடன் ஒப்பிடும் அளவிற்கும் சில சந்தர்ப்பங்களில் அதற்கும் மேம்பட்ட முடிவுகளைத் தருகின்றன.

இயந்திரம் கற்றல் செயற்கை அறிவுநுட்பத்தின் சில முக்கியக் கருத்துக்களை எடுக்கும் மற்றும் நமது சொந்த முடிவெடுக்கும் திறனை நகல்செய்யும் வகையில் வடிவமைக்கப்படும் நரம்புப் பிணையங்களுடன் உண்மையான உலகப் பிரச்சினைகளை தீர்த்து வைப்பதில் கவனம் செலுத்தும். ஆழக்கற்றல் இயந்திரம் கற்றல் கருவிகள் மற்றும் உத்திகளின் துணைக்குழு மீது இன்னும் குறுகிய கவனம் செலுத்துகிறது, மற்றும் மனித அல்லது செயற்கை சிந்தனை வேண்டும் எந்தச் சிக்கல்களையும் தீர்ப்பதற்குப் பயன்படுத்தப்படுகின்றது. முக்கியமாக ஆழக்கற்றல் ஒரு கணினி முறைக்கு நிறைய தரவுகளை ஊட்டுவதை உட்படுத்தும் இதைப் பிற தரவு பற்றிய முடிவுகளை எடுப்பதற்குப் பயன்படுத்தலாம். இயந்திரக் கற்றலில் உள்ளது போன்று இந்தத் தரவு நரம்புப் பிணையங்கள் மூலம் ஊட்டப்படுகிறது. இந்தப் பிணையங்கள் தொடர்ச்சியான பைனரி மெய்யான/தவறான கேள்விகளைக் கேட்கும், அல்லது அவற்றைக் கடக்கும் ஒவ்வொரு பிட் தரவிற்கும் ஒரு எண் மதிப்பைப் பிரித்தெடுத்து, பதில்களைப் பொறுத்து அதை வகைப்படுத்தும் தருக்க கட்டமைப்புகள் ஆகும்.

இவ்விதமான பிணையங்களை உருவாக்குவதில் ஆழக்கற்றல் செயல்பாடு கவனக்குவிப்புச் செய்வதன் காரணமாக, அவை ஆழ நரம்புப் பிணையங்களாக மாறியுள்ளன; இந்தத் தருக்கம்சார் பிணையங்கள் கூகிளின் உருவ நூலகம், அல்லது டிவிட்டரின் ட்வீட்டுகளின் / பயர்ஹவுஸ் போன்ற தரவுத் தொகுதிகளை வகைபடுத்தத் தேவையானதாகும். இவை போன்ற விரிவான தரவுத்தொகுதிகளைக் கொண்டு மற்றும் அவற்றின் வகைபடுத்தலைக் கையாளத் தக்கத் திறமையான தர்க்கப் பிணையங்களைக் கொண்டு, கணினிக்கு ஒரு உருவை எடுத்துக்கொண்டு அது மனிதர்களுக்கு எதை உருப்படுத்தம் செய்கின்றது என்பதை அதிக அளவு சாத்தியத்தோடு கூறுவது எளிதாகின்றது.

ஆழக்கற்றல் எவ்வாறு செயல்படுகிறது என்பதன் மேப்பட்ட எடுத்துக்காட்டு படங்கள் ஆகும். ஏனென்றால் படங்கள் பல வேறுபட்ட தனிமங்களைக் கொண்டுள்ளன. படங்களை நாம் விளக்கிக்கொள்வதைப் போன்று கணினி அதன் ஒரு தடக் கணக்கீடு மையப்படுத்தப்பட்ட திறனைக் கொண்டு கற்க இயலும் என்பதை நாம் புரிந்துகொள்வது எளிய

செயலன்று. ஆனால் ஆழக்கற்றலை இயந்திர குறிகைகள், கேட்டல், காணல், பேச்சு, எழுத்துசொல் போன்ற எந்தவொரு தரவு வடிவத்திலும் மனிதர்களால் பெறப்பட்டது போன்ற முடிவுகளை விளைவிக்கப் பயன்படுத்த இயலும். இதற்கு நடைமுறையான பல எடுத்துக்காட்டுகள் உள்ளன.

7. கணினிமொழியியலாரின் கவலையும் அதன் தேவையின்மையும்

கடந்த சில ஆண்டுகளில் ஆழமான கற்றல் அற்புதமான தொழில்நுட்ப முன்னேற்றங்களை அறிமுகப்படுத்தியுள்ளது என்பதில் சந்தேகம் இல்லை. ஆழக்கற்றல் மரபான இயற்கைமொழி ஆய்வைச் சவால் விடும் அளவுக்கு வளர்ந்து இயற்கைமொழி ஆய்வின் தேவையைக் குறித்து கேள்விக்கேட்கும் நிலையை எய்தியுள்ளது. ஜூன் 2015–இல் பாரிசில் பேஸ்புக் செயற்கை நுண்ணறிவு ஆராய்ச்சி ஆய்வகத்தின் துவக்க விழாவில், அதன் இயக்குனர் யான் லெசுன் ஆழமான கற்றலின் அடுத்த பெரிய முன்னேற்றம் இயற்கை மொழி புரிதலாகும் இது இயந்திரங்களுக்குத் தனிப்பட்ட சொற்களை மட்டுமன்றி முழு வாக்கியங்களையும் பத்திகளையும் புரிந்துகொள்வதற்கு திறனைக் கொடுக்கிறது. என்றார் (Mann.ng, 2015 701). இருப்பினும் மைக்கேல் ஜோர்டன் போன்ற பலரும் இயற்கை மொழி ஆய்வின் சிக்கல்களை தீர்ப்பதில் ஆழக்கற்றலின் திறனைக் குறித்துத் தங்கள் அவநம்பிக்கையை வெளிப்படுத்துகின்றனர். மானிங் இயற்கை மொழி ஆய்வு இயந்திர கற்றலின் கூடுதல் வளர்ச்சிக்கும் தொழில் பயன்பாடு பிரச்சினைகளுக்கும் மிகவும் முக்கியமாகக் கருதப்படுகிறது என்றும் எதிர்காலம் பிரகாசமானது என்றும் கருதுகின்றார் (Manning 1915 706). அவர் மொழியின் புலனறிவுசார் செயல்பாடுகளின் நுணுக்கங்களை ஆழக்கற்றல் முறை கொண்டு தீர்வுகாண இயலாது என்கின்றார். ஒரு குறிப்பிட்ட களத்தில் செய்யப்படும் ஆழக்கற்றல் முடிவுகள் வேறு களத்திற்குப் பொருத்தமாக இராது எனபதும், ஆழக்கற்றல் முடிவுகள் அவ்வபோது மாறுபடுன் என்பதும் ஆழக்கற்றல் குறித்த யுஎகைன் ஷ்மிதூபுஹ் (Jurgen Schmıahuber> 2015) எழுதியுள்ள Critique of Paper by Deep Learn ng Consp.racy" என்ற கட்டுரைக் இருங்கு கருத்தில் கொள்ளப்படவேண்டும்.

இறுதியுரை

ஆழக்கற்றல் இயந்திரக்கற்றல் போன்று இயற்கைமொழி ஆய்வுகளுக்குக் கைகொடுக்குமேயன்றி இயற்கைமொழி ஆய்வுச் செயல்பாடுகளை முற்றிலும் இடம்பெயர்க்காது. சொல் திசையன் (word vector), சொல் உட்படை (word embedding) என்பன தொடரியல் மற்றும் பொருண்மையியல் தகவல்களைக் கொண்டிருப்பதால் அவற்றை பல இயற்கைமொழி ஆய்வுகளுக்குப் பயன்படுத்தவியலும்.

துணைநூல்கள்:

wikipedia Deeplearning

Jurafsky D and Martin, J.H. (second edition) 2008. Speech and Language processing: An introduction to Natural language Processing, Computational Linguistics and speech recognition. Prentice Hall

LeCun, Y., Bengio, Y., and Hinton, G.. 2015. Deep Learning. Nature 521, 436444

Manning, Christopher D. 2015. Last Words: Computational Linguistics and Deep Learning Computational linguistics, vol. 41, no 4, pages 701-707

Marr, Bernard. What Is The Difference Between Deep Learning, Machine Learning and AI? (Downloaded from internet)

Mcenery, T. and Wilson, A. 1996. Corpus Linguistics. Edinburgh: Edinburgh University Press

Schmidhuber, Jürgen. Critique of Paper by "Deep Learning Conspiracy". Nature 521 p 436-444

தமிழில் சொற்பிழைதிருத்தியின் அவசியம்

- வல்லிபுரம் சுகந்தன்

இன்றைய கணினியுலகில் எழுத்துப்பிழைகளின்றி எழுதுவதற்கும், எழுதியவற்றில் கவனக்குறைவாக ஏதேனும் எழுத்துப்பிழைகள் விடப்பட்டிருக்கின்றனவா என்பதைச் சரிபார்ப்பதற்கும் சொற்பிழைதிருத்தியொன்றின் (spell checker) தேவை இன்றியமையாததாகும்.

சொற்பிழைதிருத்தி:

சொற்பிழை திருத்தியென்பது ஒரு கணினிச் செயலியாகும். அச்செயலியானது பொதுவாக எந்த மொழிக்குரியதோ அந்த மொழிக்குரிய சொற்கள் அனைத்தையும் ஒரு அகராதியைப்போல தன்னகத்தே சேமித்து வைத்திருக்கும். உள்ளீடு செய்யப்படும் சொற்களைத் தன் சேமிப்பில் உள்ள சொற்களுடன் சரிபார்த்து அச்சொற்கள் தன்னிடம் இல்லையாயின் அதனைப் பிழையான சொல்லாக அடையாளப்படுத்தும். இதுவே ஒரு சொற்பிழைதிருத்தியின் மிகமிக அடிப்படையான செயற்பாடாகும். அதனைத் தவிரவும் மேம்படுத்தப்பட்ட பல செயற்பாடுகளையும் அது கொண்டிருக்கும்.

ஆங்கிலமொழியிலும் தமிழ்மொழியிலும் உள்ள சொற்பிழைதிருத்திகளின் ஒப்பீடு:

ஆங்கில மொழியில் ஏற்கனவே மிகமிக மேம்படுத்தப்பட்ட சொற்பிழைதிருத்திகள் பலவுள்ளன. தமிழ்மொழியிலும் சில சொற்பிழைதிருத்திகள் ஏற்கனவே உருவாக்கப்பட்டிருப்பினும் அவற்றின் வினைத்திறன் ஆங்கிலமொழி சொற்பிழைதிருத்திகளின் வினைத்திறன் போன்று இல்லை என்பதையும், அவற்றின் பாவனை பரவலாக இல்லை என்பதையும் இங்கே கூறியாக வேண்டும். மேலும் தமிழ்மொழியிலுள்ள இச்சொற்பிழைதிருத்திகள் தமிழார்வலர்கள் எல்லோராலும் ஆங்கில மொழியிலுள்ள சொற்பிழைதிருத்திகளைப் போன்று மிக இலகுவாகப் பயன்படுத்தத்தக்க வகையிலும் இன்னமும் உருவாக்கப்படவில்லை என்பதும் கவலைக்குரியதே. அதற்கு தமிழ்மொழியின் மொழியியல்பே முக்கிய காரணமாகும்.

தமிழ்ச்சொற்பிழைதிருத்தியை உருவாக்குவதில் எதிர்ப்படும் சிக்கல்கள்:

தமிழ்மொழியில், சொற்கள் பலவிதமான உருபுகளை ஏற்கின்றன. எடுத்துக்காட்டாக அவன் என்கின்ற சொல் அவனை, அவனால், அவனொடு, அவனோடு, அவனுடன், அவனுக்கு, அவனின், அவனில், அவனது, அவன்கண், அவனுடைய என்பன போன்ற பல்வேறு உருபுகளை ஏற்பதால் இவை எல்லாவற்றையும் சேமிப்பது அவ்வளவு இலகுவான காரியம் அல்ல. அதுமட்டுமன்றி அவ்வாறு தமிழிலுள்ள அத்தனை சொற்களையும் தேடித்தேடிச் சேமித்தாலும் அவற்றைச் சேமித்து வைக்கும் தரவுத்தளம் மிகப்பெரிதாக அமைவதால், தமிழ்ச்சொற்பிழை திருத்திகளின் வினைத்திறன் குறையவும் வாய்ப்புள்ளது.

மேலும் தமிழ்மொழியில் அடுத்தடுத்து வரும் இருசொற்களில் முதற்சொல்லின் இறுதியெழுத்தையும் அதைத் தொடர்ந்து வரும் அடுத்த சொல்லின் முதலெழுத்தையும் பொறுத்து அவற்றிற்கிடையே ஒற்றெழுத்துக்களும் வரலாம். எடுத்துக்காட்டாக "அவனைக் கண்டேன்" என்று எழுதுகின்ற போது அவனை என்கின்ற உருபேற்ற சொல்லுடன் "க்" என்கின்ற ஒற்றெழுத்தும் சேர்ந்துவிடுகின்றது. அதேபோல் "அவனைப் பார்த்தேன்" என்று எழுதுகையில் "அவனை என்கின்ற உருபேற்ற சொல்லுடன் "ப்" என்கின்ற ஒற்றெழுத்து சேர்ந்துவிடுகிறது. "அவனைக்", "அவனைப்" என்கின்ற ஒற்றெழுத்துகள் சேர்ந்த சொற்களைச் சொற்பிழைதிருத்தியின் தரவுத்தளத்தினுள் சேர்க்கமுடியாது. ஏனெனில் "அவனைக் பார்த்தேன்" என்று தவறான ஒற்றெழுத்துடன் எழுதியதை இச்சொற்பிழைதிருத்தியால் அடையாளம் காணமுடியாது போய்விடும். எனவே தமிழ்ச்சொற்பிழைதிருத்தியினில் சந்திப்பிழையினைத் திருத்துவதற்குரிய சில விதிகளும் உள்ளடக்கப்படவேண்டும். அதுபோன்றே "அவனை அடித்தேன்" என்பதை "அவனையடித்தேன்" என்று புணர்த்தியெழுதுவதையும் சரிபார்ப்பதற்குரிய விதிகளையும் சொற்பிழைதிருத்தியானது கொண்டிருக்க வேண்டும்.

எழுத்துப்பிழையின்றி எழுதவேண்டியதன் அவசியம்:

எங்கள் கருத்துக்களை அல்லது தகவல்களை மற்றவர்களுக்கு எழுத்துமூலம் தெரிவிப்பதற்காகவும், தகவல்களைச் சேமித்து வைப்பதற்காகவும் அத்தகவல்களை எதிர்காலச் சந்ததியினருக்குக் கடத்துவதற்காகவும் நாம் எழுத்துமூலமான ஆவணங்களைப் பயன்படுத்துகின்றோம். அவை வெறும் செய்திகளாகவோ, இல்லை இலக்கியங்களாகவோகூட இருக்கலாம். அப்படி எம்மால் எழுத்துமூலம் சேமிக்கப்படும் அல்லது கடத்தப்படும் செய்திகள், கருத்துகள், தகவல்கள் போன்றவை எழுத்துப்பிழைகளின்றி மிகவும் திருத்தமானவையாக இருக்கவேண்டியது அத்தியாவசியமானதாகும். அவ்வாறில்லையெனின் அத்தகவல்கள் புரிந்துகொள்ளப்பட முடியாதவையாகவோ அல்லது தப்பாகவோ புரிந்துகொள்ளப்படுகின்ற அபாயமும் இருக்கிறது. இந்தத் தகவலை நாம் தொல்காப்பியத்திலும் காணக்கூடியதாக இருக்கிறது.

தமிழ்மொழியின் தொன்மைக்குச் சான்றாக இன்று எம்மிடையே எஞ்சியிருப்பவை சங்க இலக்கியங்களும், தொல்காப்பியம் என்னும் இலக்கணநூலுமாகும். தொல்காப்பியத்திற்கு முன்னர் அகத்தியம் என்னும் இலக்கணநூல் இருந்ததாகத் தொன்மக் கதைகள் இருப்பினும் அவற்றினை அறிவியல்ரீதியிலான ஆதாரமாகக் கொள்ளமுடியாது. ஆயினும் தொல்காப்பியத்தின் பல நூற்பாக்களிலும் "என்ப", "என்மனார் புலவர்" என்று தனக்கு முந்தைய அறிஞர்களைத் தொல்காப்பியர் குறிப்பிட்டிருப்பதால் தொல்காப்பியத்திற்கு முன்னும் தமிழ் இலக்கணநூல்கள் இருந்திருக்கின்றன என்பதை ஏற்றேயாக வேண்டும். தொல்காப்பியம் பொதுவாண்டிற்கு எழுநூறு ஆண்டுகள் முற்பட்டது என்பது ஏற்றுக்கொள்ளப்பட்டிருப்பதால், தமிழின் தொன்மையை நாம் ஓரளவிற்காவது உய்த்துணரக்கூடியதாக உள்ளது. அந்த ஒல்காப்புகழ்பெற்ற தொல்காப்பியம் என்கின்ற இலக்கணநூல் காரணமாகவே தமிழ்மொழி இன்றுவரை சிதையாது பாதுகாக்கப்பட்டு வருகின்றது என்று கூறினால் அது மிகையில்லை.

தொல்காப்பியத்தின் பொருளதிகாரத்தின் இறுதி இயலான மரபியலில்,

"மரபு நிலை திரிதல் செய்யுட்கு இல்லை
மரபு வழிப்பட்ட சொல்லினானே"

"மரபு நிலை திரியின் பிறிது பிறிது ஆகும்"

என்கின்ற நூற்பாக்களின் மூலம் சொற்களின் மரபு செய்யுட்களில் (இக்காலத்தில் அது எழுத்துவடிவில் உள்ள இலக்கியங்கள், கட்டுரைகள் செய்திகள் போன்றவற்றிற்கும் பொருந்தும் எனக் கொள்ளலாம்) பேணப்படவேண்டுமெனவும், அவ்வாறு இல்லையெனின் அவை சிதைந்து வேறாகிவிடுமெனவும் கூறப்பட்டிருக்கிறது. இது சொற்களில் நேரும் எழுத்துப்பிழைகட்கும் பொருந்துகின்றது என்பதும் குறிப்பிடத்தக்கது.

எனவே கணினியுலகில், தமிழ்மொழியானது சிதைந்துபோகாமல் தொடர்ந்து உரியமுறையில் பேணப்படுவதற்கு, கணினித்தமிழில் ஏற்படும் சொற்பிழைகளைத் திருத்துவதற்கான சொற்பிழைதிருத்தியொன்று மிகமிக இன்றியமையாததாகும்.

மின்னணுக்கருவிகள், மென்பொருள்களில் தமிழை நிலைப்படுத்த

முனைவர் மா. தமிழ்ப்பரிதி
உதவிப்பேராசிரியர், இதழியல் – மக்கள் தொடர்பியல்துறை,
பெரியார் பல்கலைக்கழகம், சேலம் – 636011 தமிழ்நாடு
மின்மடல்: tparithi@gmail.com
உலாபேசி: 91+7299397766

இவ்வுலகின் தொன்மையான மொழிகளுள் ஒன்றாக இன்றும் மக்களால் ஆளப்படும் மொழியாக தமிழ் பெரும்பாலான நாடுகளில் புழக்கத்தில் உள்ளது. காலத்தின் தேவையாக முகிழ்ந்திருக்கும் மின்னணுக்கருவிகளிலும் மென்பொருள்களிலும் இணையத்திலும் தமிழின் பயன்பாட்டினை மேம்படுத்துவது மிகத்தேவையான ஒன்றாகும். மின்னணுக்கருவிகளில் தமிழிற்கான ஒத்திசைவுகளை மிக நுண்ணிய முறையில் ஆராய்வதும் ஒத்திசைவில் காணப்பெறும் இடைவெளிகளைக் களைவதும் உடன் நிகழ வேண்டிய ஒன்றாகும். மின்னணுக்கருவிகளில் தமிழ் நிலைபெற, அதற்குரிய செயற்திட்டங்களை நாம் முன்னெடுக்க வேண்டியது காலத்தின் தேவையாகும்.

இணையத்தில் இந்திய மொழிகளில் தமிழில் பயன்பாடு முதலிடத்தில் உள்ளதை (42%) கூகிளின் அண்மைய ஆய்வுகள் வெளிப்படுத்தியுள்ளன. உலகம் பரவி வாழும் கோடிக்கணக்கான தமிழர்களின் வாழ்வியல் அடையாளமாக விளங்கி வரும் தமிழை அனைத்து வகையான மின்னணுக்கருவிகளுக்கும், மென்பொருள்களுக்கும் ஒத்திசைவானதாகவும், பொருத்தப்பாடு உடையதாகவும் மாற்ற பன்னாட்டுத் தன்மையிலான கொள்கைகளை வகுப்பதும், அதனை செயற்படுத்துவதும் உடனடித்தேவையாகும்.

தமிழின் உள்ளீடு / வெளியீடு ஆகியனவற்றைச் செயற்படுத்துவதற்கான கருவி அமைப்பு, நிரலாக்கம், மென்பொருட்களின் உருவாக்கம், தரப்பாட்டுக் கொள்கை, கண்காணிப்பு முறைகள் ஆகியனவற்றை உருவாக்க வேண்டியது உடனடித்தேவையாகும்.

அனைத்து மின்னணுக்கருவிகளிலும் தமிழ்

அனைத்து மின்னணுக்கருவிகளிலும் தமிழிற்கான ஒத்திசைவை ஏற்படுத்துவதை அரசின் உறுதியான கொள்கை முடிவால் மட்டுமே செயற்படுத்த முடியும். எனவே, எந்த மின்னணுக்கருவியாக இருந்தாலும் அவற்றில் தமிழின் ஒத்திசைவிற்கு அரசிடன் சான்றிதழ் பெற வேண்டும். தமிழுக்கு ஒத்திசைவுச்சான்றிதழினைப் பெறும் மின்னணுக்கருவிகள் மட்டுமே விற்பனையுரிமை வழங்கப்பெற வேண்டும்.

அனைத்து மின்னணுக்கருவிகளிலும் மென்பொருள்களிலும் தமிழை நிலைப்படுத்துவதற்கான செயல்திட்டத்தில் கணித்தமிழ் ஆர்வலர்களும், களச்செயற்பாட்டாளர்களும் இணைந்து செயல்பட்டு தமிழை மின்னணுக்கருவிகளில் நிலைப்படுத்த உழைக்க வேண்டும்.

தமிழ் ஆட்சி மொழியாகவும், தேசிய மொழியாகவும், தொடர்பு மொழியாகவும் உள்ள நாடுகளில்

மின்னணுக்கருவிகளில் தமிழை நிலைபெறச்செய்வதை தமிழ்க்கணிமை ஆர்வலர்கள் கண்காணித்து அரசுக்கு உறுதுணையாகச் செயற்பட வேண்டும்.

இணையதளங்களில் தமிழ்

தமிழ்நாட்டரசின் இணையதளங்களில் தமிழ் ஒன்பது விழுகாட்டிற்கும் குறைவாக உள்ளதை ஆய்வுகள் வெளிப்படுத்தியுள்ளன. தமிழ் ஆட்சிமொழிச்சட்டம் 1956, தகவலறியும் உரிமைச்சட்டம் 2005 குறித்த உள்ளுணரலில் உள்ள இடைவெளியினை நம்மால் அறிய முடிகின்றது. தமிழக அரசின் இணைய தளங்கள் தமிழில் இருக்க வேண்டும், முகப்புப்பக்கம் உறுதியாக தமிழில் இருக்க வேண்டும் போன்ற அறிவிப்புகளை அரசு உடன் வெளியிட்டு அதனை முனைப்புடன் செயற்படுத்திக் கண்காணிக்க வேண்டும். மக்கள் தொடர்பிற்கும் நல்லுறவிற்கும் இணையங்கள் அடிப்படையாக விளங்குவதால் இணைய உள்ளடக்கங்களில் தமிழ் வழித்தொடர்பாடலை உறுதி செய்வது அரசின் அடிப்படைக் கடமையாகும். தமிழக அரசின் அதிகார வழிப்பட்ட அனைத்து இணையதளங்களும் கண்டிப்பாக ஒருங்குகுறித்தமிழில் இருக்க வேண்டும் என அரசு, தமிழ் ஆட்சிமொழிச்சட்டத்தில் உடனடியாக திருத்தம் செய்ய வேண்டும். தமிழக அரசின் அதிகார வழிப்பட்ட அனைத்து இணையதளங்களையும் கண்காணிப்பதன் பொருட்டு தமிழ் இணையதள ஒழுங்கமைப்பு ஆணையம் ஒன்று ஏற்படுத்தப்பட்டு அதன்வழி இணையத்தில் தமிழ் உள்ளடக்கங்கள் கண்காணிக்கப்பெற வேண்டும்.

தமிழக அரசின் இணையதளங்களில் முதன்மைப்பக்கம் தமிழில் இருக்க வேண்டும். குறிப்பிட்ட கால இடைவெளியில் தமிழக அரசின் இணையதளங்களில் தமிழ் மின்னாட்சியை அரசு கண்காணிக்க வேண்டும். தமிழக அரசு ஊழியர்களுக்கு ஆட்சித்தமிழ் செயற்பாட்டில் கணித்தமிழின் தேவை குறித்த உள்ளுணர்வை மேம்படுத்த வேண்டும். அரசின் இணையதளங்கள், கையடக்கக்கருவிகளில் பார்க்கும் வகையில் ஒத்திசைவுடன் அமைதல் வேண்டும். நாடும் அதன் மக்களும் தன் தன் ஆட்சிமொழி குறித்து குறித்து நாம் சிந்தித்து அதனைக் காக்க வேண்டிய நேரமிது.

தமிழில் இயங்குதளங்கள், மென்பொருள்களின் தமிழ்ப்பதிப்பு

தமிழக அரசு கொள்முதல் செய்யும் இயங்குதளங்கள், மென்பொருட்கள் தமிழ்ப்பதிப்பைக் கொண்டதாக அமைய வேண்டும். இயங்குதளங்கள், மென்பொருட்களில் தமிழின் ஒத்திசைவிற்கு தமிழக அரசு சான்றளிப்புச்செய்ய வேண்டும்.

தமிழ்நாடு மின்னணுக்கழகம் (வரை.) கொள்முதல் செய்யும் மென்பொருட்கள், இயங்குதளங்கள், மின்னணுக்கருவிகளில் தமிழின் ஒத்திசைவை அடிப்படையாக்க வேண்டும்.

தமிழ் ஒருங்குகுறி, தமிழ் விசைப்பலகை

மாணவர்களுக்கு வழங்கப்பெறும் மடிக்கணிப்பொறிகளில் தமிழ்99 விசைப்பலகையைத் தமிழக அரசு அச்சிட்டு 2017 கல்வியாண்டில் 5.3 இலக்கம் மாணவர்களுக்கு வழங்க உள்ளமை கணித்தமிழ் வளர்ச்சியில் ஒரு முதன்மையான திருப்புமுனையாகும்.

ஆப்பிள் நிறுவனத்தின் கணிப்பொறிகளிலும், அலைபேசிகளிலும் தமிழ் இயல்பான உள்ளீட்டு மொழியாக செயல்பாட்டில் உள்ளது. எச்டிசி அலைபேசியில் தமிழினை இயல்பான உள்ளீட்டு மொழியாகப் பயன்படுத்த இயலும். அலைபேசிகளின் ஆண்டிராய்டு இயங்குதளம் தமிழிற்கான ஒத்திவைக் கொண்டிருந்தாலும், ஒருவர் தமிழ் உள்ளீட்டிற்காக ஏதேனும் மென்பொருளினைப் பதிவிறக்குவது அதனை நிறுவுவது என்பதும் அலைபேசிகளில் தமிழின் இருத்தல் தொடர்பான இடைவெளிகளில் ஒன்றாகும். அலைபேசி நிறுவனங்களுக்கு வழிகாட்டல் செய்து தமிழ் 99 விசைப்பலகையை எந்தொரு மென்பொருளையும் தரவிறக்காமல், நிறுவாமல் தமிழ் 99 உள்ளீடு செயற்படுத்தப்பெற வேண்டும்.

தமிழ் 99 என்னும் விசைப்பலகை முறையினை அரசு செந்தரப்படுத்தி இருப்பினும் பெரும்பாலோனோர் இன்னமும் எழுத்துப்பெயர்ப்பு முறையிலான தமிழ் உள்ளீட்டினையேப் பின்பற்றி வரும் இன்றைய நிலையில், தமிழ் 99 விசைப்பலகையினைப் பரவலாக்குவதன் பொருட்டு தமிழ்த்தட்டச்சு தேர்வில் தமிழ்99 விசைப்பலகை முறையிலான தட்டச்சு முறையைச் செயல்படுத்திட வேண்டும். தமிழ்99 விசைப்பலகை தட்டச்சினை எளிதாகக் கற்பதற்கு உரிய விளையாட்டுகளை உருவாக்குதல், கையேடுகளை வழங்குதல் போன்றவற்றிற்கான வழிமுறைகளை அரசும், தமிழ்க்கணிமை ஆர்வலர்களும் முன்னெடுக்க வேண்டும்.

தமிழக அரசு அலுவலகங்களில் தமிழ்99 விசைப்பலகைகளே இனிமேல் கொள்முதல் செய்யப்பெறுதல் வேண்டும். தமிழக அரசு ஊழியர்களுக்கும் மாணவர்களுக்கும் கணித்தமிழ்ப்பயிற்சிகள் அளிக்கப்பெற வேண்டும். தமிழ் 99 விசைப்பலகை விண்டோசு இயங்குதளத்தில் சேர்க்கப்பெற வேண்டும்.

கணித்தமிழ்க்கல்வி

தமிழகத்தின் அனைத்துப் பல்கலைக்கழகங்களிலும் கணித்தமிழுக்கான கல்வி, ஆய்வுத்துறைகளை நிலைப்படுத்திட கணித்தமிழ்த்துறைகள் ஏற்படுத்தப்பெற வேண்டும். கணித்தமிழ் ஆய்வுகளை உலகத்தரத்தில் முன்னெடுக்க ஆய்வுத்தகைமைகள் வழங்களப்பெற வேண்டும். தமிழில் மொழிக்கணிமையியல் சிலரின் ஆர்வத்தால் இயங்கும் ஒன்றாக உள்ளது. இதை மாற்ற தமிழக அரசின் உறுதுணையும், உயர்கல்வி நிறுவனங்களின் உள்ளார்ந்த ஈடுபாடும் மிகத்தேவையான ஒன்றாகும்.

தமிழ் விக்கிப்பீடியா போன்ற பொதுவள ஊடகத்தில் தமிழ் உள்ளடங்கங்களை மேம்படுத்துவது குறித்து மாணவர்களுக்கும், பொதுமக்களுக்கும் தொடர் பயிற்சிகள் அளிக்கப்பெற வேண்டும்.

கணித்தமிழ் காலத்தை வெல்ல

அனைத்து மின்னணுக்கருவிகளிலும் தமிழை உறுதி செய்ய வேண்டியது அரசின் கடமையாகும். மின்னணுக்கருவிகளில் தமிழிற்கான ஒத்திசைவை ஏற்படுத்துவது மட்டுமே தமிழ் மின்னாட்சியில்ஃ மின்னாளுகையில் தமிழை நிலைப்படுத்தும். அனைத்துவகை மென்பொருள்களிலும் தமிழ்ப்பதிப்புகளை அரசு உறுதி செய்ய வேண்டும். கணித்தமிழின் நிலைப்பேறாக்கம் அரசு, அரசு நிறுவனங்கள், கணித்தமிழ் அறிஞர்கள்ஃ ஆர்வலர்கள் ஆகியோரின் கூட்டுச்செயற்பாட்டால் நிகழ வேண்டிய ஒன்றாகும்.

கணித்தமிழ் வளர்ப்போம், காலத்தை வெல்வோம்.

பார்வைகள்

1. https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf

2. Maa.Thamizhpparithi, Portals of Government of Tamil Nadu: Communication in Tamil Present and Futurology, Aayvagam an International Journal of Multidisciplinary Research, Vol:1, Issue:7, August 2013, Pg.1-14-17

3. http://tamil.thehindu.com/tamilnadu/இணைய தமிழ் அகராதி தொகுப்பில் ஒரு முன்முயற்சி/article9255822.ece

16 வது தமிழிணைய மாநாட்டுக்கு நன்கொடை அளித்த நல்லுள்ளங்கள்

- திரு. ஆண்டி கிரி Mr. Andi Giri
- திரு. குமார் குமரப்பன் Mr. Kumar Kumarappan
- முனைவர். வாசு அரங்கநாதன் Dr. Vasu Renganathan
- முனைவர். ந. பரமேசுவரன் Dr. N. Parmeswaran
- மரு. திருமாவளவன் + உமா திருமாவளவன் Dr. Thirumavalavan + Uma Thirumavalavan
- முனைவர்.இரா. செல்வராசு Dr.R Selvaraj
- மரு. சோமசுந்தரம் இளங்கோவன் Dr. Somasundaram Ilangovan
- திரு. தில்லை குமரன் Mr. Thillai Kumaran
- திரு. அருள் வீரப்பன் – Mr. Arul Veerappan



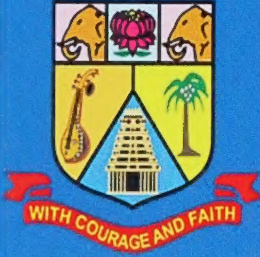
16th TIC – Toronto- List of Accepted Papers			
No.	**Authors**	**Title (shortened)**	**Country**
1	Shanti Ramalinggam	The role of VLE Frog in assisting students, teacher and parents in Mlearning and usage of ICT tool such as smartphones and computational devices in school curriculum.	Malaysia
2	Sivakumaran	கற்பித்தலில் தரவக மொழியியலின் பங்கு	Singapore
3	Ram Krishna Pandey and A.G. Ramakrishnan	Tamil Document-Image Super-Resolution for improved OCR recognition	India
4	Brenda Beck	Digital Archive for Tamil Texts	Canada
5	S. Johanan Joysingh, M. Nanmalar, G. Anushiya Rachel, V. Sherlin Solomi, V. Dhanalakshmi, P. Vijayalakshmi, T. Nagarajan	Development of a Speech-Enabled Interactive Enquiry System in Tamil for Agriculture	India
6	Vasu Renganathan	Challenges of Machine Learning with Tamil Texts from Ancient to Modern Tamil	USA
7	A.Madhavaraj and A.G. Ramakrishnan	Deep neural network based medium vocabulary continuous speech recognition system for Tamil	India
8	Vasanthan Thirunavukkarasu, Jonathan P. Evans, Sankar Raman, Sachit Mahajan, Mrinal Kanti Baowaly, Priyadharsini Karuppuswamy, and Sailesh Rajasekaran	Quantifying shifts in language use among internet-using Tamil speakers	Taiwan
9	மு. மயூரன்	இலங்கையில் அரசகருமமொழிகள் நடைமுறையாக்கத்தின் ஒரு பகுதியான தமிழ்மொழி நடைமுறையாக்கத்தில் தகவற் தொழிநுட்பத்தின் வகிபாகம்	Sri Lanka
10	முனைவர் விசயராணி	பார்வை மாற்றுத்திறனாளிகளின் தமிழ்மென்பொருள் பயன்பாட்டில் தேவைகள், சிக்கல்கள், தீர்வுகள்	India
11	Ku. Kalyanasundaram and Vasu Renganathan	Digitization, Distribution and Synthesizing Tamil Texts: Challenges of taking Madurai Project to its next step	Switzerland/USA
12	Ram Kallapiran [Pudhuyugan]	Is the use of Digital Archiving a critical need today in preserving a classical language?	UK
13	Vamuse Andavar	Computer aided case marker for Thirukkural	India
14	Shalini and Vijayalakshmi	Collborative and Interative Video Quiz in Tamil using Computational Offloading.	India
15	Sairam Jayaraman and Muruganandam Sundarrajan	"2016ம் ஆண்டு தமிழ்நாடு சட்டப்பேரவை தேர்தலும், தமிழக இளைஞர்களின் அரசியல் சார்ந்த சமூக இணையதளப் பயன்பாடும்", not "தமிழ்நாடு சட்டப்பேரவை"	India
16	K.Mrinalini,. G. Anushiya Rachel, T. Nagarajan, and Vijayalaksjmi	Sentence-medial pause identification for Tamil synthesis system	India
17	A.G.Ramakrishnan and K.V. Vijay Girish	Enhancement of noisy Tamil speech for improved quality of perception for the hearing impaired	India

18	Shahul Hameed M M (Shah)	Engaging Augmented Reality and Collaborating With Learners to Inspire and Maximize Learning of Tamil Language	Singapore
19	S. Pushpa Rani	இலக்கணப் பிழைகளின்றி தமிழ் எழுதிட எட்மோடோ) Edmodo) வழி மெய்நிகர் கற்றல் கற்பித்தல் அணுகுமுறை	Malaysia
20	கருணாகரன் கணேசன், கிரேச்குமார் ராமராசு, அருண்ராம் ஆத்மசரன், முத்து அண்ணாமலை.	எழில் - பொது பயன்பாட்டிற்கும், வெளியீடு நோக்கிய சவால்களும்	USA
21	திரு சிவா சுப்பிரமணியன், திரு செந்தில்	தமிழ் APIகள் மூலம் இணைப்பில் இல்லா இணையதளங்களை) Offline Websites) தமிழில் உருவாக்குதல், அவற்றின் முக்கியத்துவம் மற்றும் பயன்கள் - ஓர் ஆய்வு	UK
22	திரு. செல்வமுரளி	தமிழின் பெருந்தரவகத் தரவுகள் தேவையும், பயன்பாடும்	India
23	Muthiah Annamalai and T. Shrinivasan	Tamil Open-Source Lanscape – Opportunities and Challenges	USA/India
24	செ. ஜெயபாலன், இ. ரோகினி	தமிழ் மொழி மூலமான இணைய வணிக பொருளாதார சேவைகள் மையங்களை பிரதேசங்கள் தோறும் உருவாக்கி நடைமுறைப்படுத்துவதில் காணப்படும் சவால்களும் அதனை வெற்றிகொள்ள வேண்டிய தந்திரோபாயங்களும் இலங்கையின் மட்டக்களப்பு மாவட்டம் தொடர்பான ஆய்வு.	Sri Lanka
25	இ. நற்கீரன்	தமிழ்ச் சூழலில் திறந்த இணைப்புத் தரவுக்கான மெய்ப்பொருளிய உருவாக்கம் நோக்கி	Canada
26	சிவபாலன் திருச்செல்வம், மேனேஸ் , ரூபிணி தியாகராஜன்	தமிழ் கற்றல் கற்பித்தலில் 21ஆம் நூற்றாண்டுத் தகவல் தொழில்நுட்ப மதிப்பீடு : குயிசிஸ் (Quizizz)	Malaysia
27	Withdrawn		~~Singapore~~
28	கஸ்தூரி இராமலிங்கம்	ஊடாடல், நகர்ப்படங்கள் கலந்த மின்னூல்கள் வழிக்குழந்தைகளுக்கான தமிழ்க்கல்வி	Malaysia
29	மா . ஜெயகானந்தன், கு . வினுஜனன், செ . ஜெயபாலன்	அடுத்த தலைமுறைக்கான தமிழ் மொழி நூல்களின் ஒருங்கிணைக்கப்பட்ட தேடு	Srilanka
30	R.Kalaiselvan 1 Dr.T.Mala 2 Shri Vindhya	CROSS LINGUAL PERSONALIZED TRAVELRECOMMENDATION USING LOCATION BASED SOCIAL NETWORKS	India
31	Gokul Kumar.M, Keerthana.V, Hemanandhini.S, and Dr.T.Mala	PADAM INAITHAL-WORDS MATCHING WITH IMAGES GAME	India
32	Withdrawn		~~Singapore~~
33	M. Antony Robert Raj, S. Abirami	Offline Tamil Handwritten Character Recognition: Challenges	India
34	Elango Cheran	Prefix Trees (Tries) for Tamil Language Processing	USA

OUR SPONSORS



Canadian Tamil Congress
கனடியத் தமிழர் பேரவை
Congrès Tamoul Canadien
"Voice of Tamil Canadians"

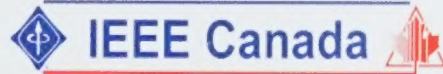


அண்ணாமலை கனடா வளாகம்
Annamalai Canada Campus
1240 Ellesmere Rd. Scarborough. ON. Canada. TP: 416-777-2822



UNIVERSITY OF
WATERLOO

CENTRE FOR PATTERN ANALYSIS
& MACHINE INTELLIGENCE





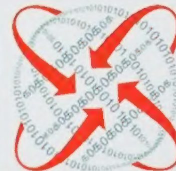


UNIVERSITY OF
TORONTO

2017

உத்தமம்

INFITT



உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம்

International Forum for Information Technology in Tamil